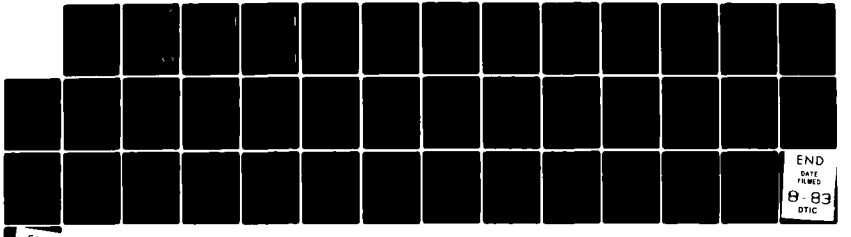
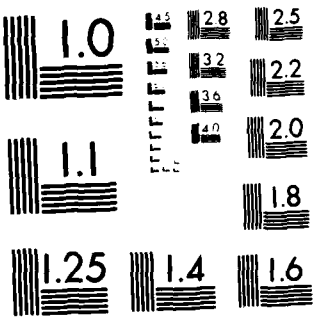
AD-A130 406 DYNAMIC PROPERTIES OF ELECTRONIC TRAPPING CENTERS AT THE SI-SIO2 INTERFACE(U) XEROX PALO ALTO RESEARCH CENTER CA GENERAL SCIENCES LAB N M JOHNSON MAY 83 1/1

UNCLASSIFIED DAAK20-81-C-0406 F/G 20/12 NL



END DATE FILMED 8-83 DTIC

END DATE FILMED 8-83 DTIC



(12)

DYNAMIC PROPERTIES OF ELECTRONIC TRAPPING CENTERS AT THE Si-SiO_2 INTERFACE

FINAL REPORT

N. M. JOHNSON

May 1983

Electronic Materials Research Division
U. S. Army Electronics Technology & Devices Laboratory
Fort Monmouth, New Jersey 07703

Contract No. DAAK20-81-C-0406

Xerox Palo Alto Research Center
Palo Alto, California 94304

(unclassified)

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Final Report	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Dynamic Properties of Electronic Trapping Centers at the Si-SiO_2 Interface		5. TYPE OF REPORT & PERIOD COVERED Final Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) N. M. Johnson		8. CONTRACT OR GRANT NUMBER(s) DAAK20-81-C-0406
9. PERFORMING ORGANIZATION NAME AND ADDRESS General Sciences Laboratory Xerox Palo Alto Research Center Palo Alto, CA 94304		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS --
11. CONTROLLING OFFICE NAME AND ADDRESS Contracting Offices Naval Research Laboratory Washington, DC 20375 Attn: Code 1231.AB		12. REPORT DATE May 1983 13. NUMBER OF PAGES 25
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) DCASMA San Francisco 1250 Bayhill Drive San Bruno, CA 94066		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution Unlimited

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Distribution Unlimited

18. SUPPLEMENTARY NOTES

--

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Silicon Dioxide
SiO_2-Si Interface
Electronic Defects
Paramagnetic Defects
Amphoteric Defects
Deep-Level Transient Spectroscopy

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

See reverse side.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Abstract

On unannealed, thermally oxidized silicon, electron spin resonance reveals an oriented interface defect which is termed the P_b center and identified as the trivalent silicon defect. Deep-level transient spectroscopy (DLTS) reveals two broad characteristic peaks in the interface-state distribution: one ~0.3 eV above the silicon valence-band maximum, and a second ~0.25 eV below the conduction band. Isochronal anneals of oxidized silicon, coated with aluminum, show that the spin density and the densities of the two DLTS peaks have the same annealing kinetics. On large-area, Al-gated capacitors the spin density can be modulated with an applied voltage; sweeping the silicon bandgap at the interface through the Fermi level reveals that the spin density is approximately constant over the central region of the bandgap but decreases near the band edges. The variation of the spin density with gate voltage identifies an amphoteric center with both electronic transitions in the bandgap. Both the annealing behavior and the voltage dependence of the $P_{(b)}$ center support the conclusion that these transitions correspond to the two characteristic peaks in the interface-state distribution. The ~0.6 eV separation of the peaks is the effective correlation energy of the dangling orbital on a trivalent silicon defect at the $Si-SiO_2$ interface. The similarity between the disordered interface and amorphous silicon and the effects of a correlation energy on deep-level measurements are discussed.

TABLE OF CONTENTS

Abstract

On unannealed, thermally oxidized silicon, electron spin resonance reveals an oriented interface defect which is termed the P_b center and identified as the trivalent silicon defect. Deep-level transient spectroscopy (DLTS) reveals two broad characteristic peaks in the interface-state distribution: one ~0.3 eV above the silicon valence-band maximum, and a second ~0.25 eV below the conduction band. Isochronal anneals of oxidized silicon, coated with aluminum, show that the spin density and the densities of the two DLTS peaks have the same annealing kinetics. On large-area, Al-gated capacitors the spin density can be modulated with an applied voltage; sweeping the silicon bandgap at the interface through the Fermi level reveals that the spin density is approximately constant over the central region of the bandgap but decreases near the band edges. The variation of the spin density with gate voltage identifies an amphoteric center with both electronic transitions in the bandgap. Both the annealing behavior and the voltage dependence of the P_b center support the conclusion that these transitions correspond to the two characteristic peaks in the interface-state distribution. The ~0.6 eV separation of the peaks is the effective correlation energy of the dangling orbital on a trivalent silicon defect at the $Si-SiO_2$ interface. The similarity between the disordered interface and amorphous silicon and the effects of a correlation energy on deep-level measurements are discussed.

I. INTRODUCTION

Thermal oxidation of single-crystal silicon produces characteristic electronic defects at the $Si-SiO_2$ interface. Electrical measurements on metal-oxide-silicon (MOS) devices reveal a continuous distribution of interface states that extends throughout the silicon bandgap. For interfaces obtained by thermal oxidation without a subsequent anneal, a broad peak in the interface-state distribution centered ~0.3 eV above the silicon valence-band maximum (i.e., $E_V + 0.3$ eV) has been observed.[1-3] Electron spin resonance (ESR) reveals a paramagnetic defect, designated the P_b center, at the interface which has been identified as a trivalent silicon defect bonded to three silicon atoms, with the nonbonding orbital aligned along the (111) directions.[4] A direct relationship between the above manifestations of interface defects has been proposed[1-3] based on theoretical studies of the density of states associated with the trivalent silicon defect.[5] Modulation of the P_b spin signal with an applied electric field was first demonstrated[6] by varying the gate voltage on an MOS capacitor during ESR measurements and subsequently confirmed[7] by using the corona charging method to vary the field.

This report presents a detailed comparison of electrical and spin-resonance measurements of the P_b center and deep electronic levels at the interface. Results from ESR, capacitance–voltage (C–V) measurements, and deep–level transient spectroscopy (DLTS) are shown to be consistent with an amphoteric model for the P_b center in which the characteristic interface states correspond to the electronic levels of the dangling orbital on a trivalent silicon defect. The similarity between the disordered interface and amorphous silicon is discussed, and the implications of the amphoteric nature of interface defects on deep–level measurements are examined. In the appendix is presented a critical

review of the measurement of semiconductor–insulator interface states by constant–capacitance, deep–level transient spectroscopy (CC–DLTS).

II. EXPERIMENTAL RESULTS

A. Interface–State Distribution

Results from DLTS measurements of interface states across virtually the entire silicon bandgap are shown in Fig. 1. Identically processed n–type and p–type MOS capacitors were used to determine the density of states in the upper and lower halves, respectively, of the bandgap. The devices were fabricated on (111) – oriented epitaxial silicon. The oxide was grown in dry O_2 at 1000°C to a thickness of 110 nm. Aluminum films were deposited from a flash – evaporation source to form gate electrodes and back (Ohmic) contacts, without introducing radiation damage into these unannealed test devices. The current – transient mode of measurement[8] was used to implement a high emission rate window ($e_0 = 1 \times 10^4\ s^{-1}$) in order to detect emission from interface states near the band edges at temperatures down to 30 K. The analysis assumed constant capture cross sections for electrons (n–type) and holes (p–type) based on previous measurements on similarly prepared devices.[2,9] Near each band edge a peak is resolvable from the band tail of interface states; the dashed line segments indicate the region near midgap where the DLTS majority – carrier analysis is generally inaccurate as discussed in Ref. 8. In the present study, no attempt was made to deconvolve the band tails from the peaks, hence, the band – tail distributions are only approximate since it is anticipated that their capture cross sections will differ from those of the peaks. These peaks, which are characteristic of thermally oxidized silicon, appear at approximately $E_V + 0.3$ eV and $E_C - 0.25$ eV and are of equal density within experimental uncertainty; their

separation is ~0.6 eV. Both peaks were removed by annealing in forming gas at 450°C (30 min).

B. Isochronal Anneals

Isochronal anneals were used to evaluate the relative annealing kinetics of the P_b spin center and the characteristic peak at $E_V + 0.3$ eV in the interface-state distribution. The samples consisted of epitaxial silicon on (111)-oriented silicon substrates, with semiconducting n-type epitaxial layers. For electrical measurements the substrates were degenerately doped for n^+ conductivity; for ESR undoped high-resistivity substrates were used to minimize loading of the microwave cavity. The different silicon wafers were simultaneously processed to fabricate MOS test structures. A thermal oxide was grown in dry O_2 at 1000°C to a thickness of 110 nm, and an Al film, 300-nm thick, was vacuum evaporated onto the oxide layer. Isochronal anneals of 15 min duration were performed under vacuum. The specimens were annealed in pairs, with one sample for C-V measurements and the other for ESR. The density of the $E_V + 0.3$ eV peak was determined on the n-type capacitors by the quasi-static (i.e., high-low) CV technique.[10] For spin-density measurements, the Al film was removed before ESR in order to minimize cavity loading.

In Fig. 2 the characteristic peak density and the P_b spin density, both normalized to their maximum (unannealed) values, are plotted together for the series of isochronal anneals. A maximum spin density of $\sim 1.5 \times 10^{12}$ cm^{-2} was determined by numerically integrating the ESR signal from a stack of samples and comparing with a weak pitch standard.[3] The maximum density of interface states at the $E_V + 0.3$ eV peak was $\sim 1.6 \times 10^{13}$ $eV^{-1}cm^{-2}$. These results reveal that the P_b center and the characteristic deep levels anneal identically. A check of the annealing behavior of the $E_V + 0.3$ eV peak on p-type MOS capacitors

yielded similar results. In addition, the $E_C - 0.25$ eV peak anneals similarly to the lower peak and spin center, although this was more difficult to establish due to its close proximity to a large residual conduction band tail. An essential factor in achieving a meaningful correlation between these two manifestations of interface defects is the identical chemical and thermal treatments which the different specimens received in preparation for the measurements. This is consistent with the generally held view that annealing with aluminum involves process-dependent residual water in the oxide and/or at the oxide-metal interface.[3] The simultaneous processing of the different wafers in the present study insured that this factor had an identical influence on all of the specimens.

C. Electric-Field Modulation of the P_b Spin Density

Large-area, Al-gated MOS capacitors were used to electric-field modulate the P_b spin density during ESR measurements. Czochralski-grown, p-type (30 Ω cm), (111)-oriented silicon wafers were oxidized in dry O_2 at 1000°C. A high dose of boron was implanted into the back side of the wafers and activated during oxidation to provide a p^+ layer for Ohmic contact after metallization. Capacitors with an Al-gate area of 0.6 cm^2 displayed negligible leakage under strong accumulation and inversion biases, and the flatband voltage was approximately -12 V. For ESR measurements, the back metal and underlying p^+ layer were selectively etched to define a small bonding pad which minimally loaded the microwave cavity.

The dependence of the P_b spin signal on gate voltage is shown in Fig. 3. Each point is the relative peak-to-peak amplitude of the P_b (derivative) absorption spectrum which was recorded at a nonsaturating microwave power and averaged over ten or more scans. In addition, the relative amplitudes have been corrected for any small background signal by subtracting the signal from

the same sample after a 300°C anneal. The P_b spin signal decreases for either high negative or positive gate voltages which coincidentally correspond to strong accumulation and inversion, respectively, of the MOS capacitor. The signal is featureless within experimental error for gate voltages between the onset of strong accumulation and inversion, that is, for gate voltages roughly in the range from −20 to 0 volts. This range corresponds to the Fermi level intersecting the interface over the central region of the silicon bandgap. For a given gate voltage this intersection occurs over a distribution of energies due to surface potential fluctuations,[10] the standard deviation of which is largest in depletion and estimated to be $\leq$ 0.14 eV. The dependence of the ESR signal on V_G nearly mirrors the electric-field induced free-carrier concentration at the interface, so rapid relaxation of interface states by free carriers can (and does) occur which could cause an apparent loss of signal through line broadening. However, voltage invariance of the line shape and saturation measurements of the spin-lattice relaxation time demonstrated that the relaxation rates were too low. The reduction in signal amplitude is thus due to a change in defect occupancy induced by the change in surface potential.

The declines of the spin density at each end of the voltage range identify an amphoteric defect center. At high negative voltages, the Fermi level intersects the interface near the silicon valence-band maximum, and the center is diamagnetic; in the intermediate voltage range, the center contains an unpaired electron and yields an ESR signal; for positive gate voltages, the Fermi level approaches the conduction band, and the center acquires a second electron which pairs with the first to again produce a diamagnetic site. Similar results have been obtained on n-type MOS capacitors, where it was also determined that the maximum spin density on a metallized capacitor is the same as that

measured with the electrodes removed; that is, in the absence of an applied voltage, the Fermi level is pinned by the defect states.

III. AMPHOTERIC DEFECT MODEL

There are striking similarities between the Si–SiO_2 interface and hydrogenated amorphous silicon, a–Si:H. The density of gap states in a–Si:H, shown schematically in Fig. 4, is composed of band tails of localized states (strained bonds) and a single dominant electrically active defect, the trivalent silicon atom, which has two levels in the mobility gap.[11-14] These levels correspond to single and double electron occupancy of the dangling orbital and are separated by an effective correlation energy $U_0 \approx 0.4$ eV.[13,15] The effective correlation energy is the difference between the energies required to add an electron to a singly occupied orbital and to an unoccupied site; it is the sum of two terms, a Coulombic repulsion and a compensating lattice relaxation energy. The electron occupancy of the orbital varies with the Fermi energy as described in Sect. II.C, with the singly occupied orbital charge neutral.

The amphoteric defect model for a-Si:H is directly applicable to the P_b center at the Si–SiO_2 interface, which has also been identified as a trivalent silicon defect.[4,11] The amphoteric nature of the center is demonstrated in Fig. 3, which also establishes that both electronic transitions of the center are in the silicon bandgap. Correlating the P_b center with the characteristic DLTS peaks (Figs. 1 and 2) supports the assignments of the 0→1 electron transition of the amphoteric center to the $E_V + 0.3$ eV peak and the 1→2 electron transition to the $E_C - 0.25$ eV peak. Then the ~0.6 eV separation of the peaks corresponds to the effective correlation energy of the trivalent silicon defect at the Si–SiO_2 interface.[16] In comparison with a-Si:H, the larger correlation energy for a Si dangling orbital at the Si-SiO_2 interface primarily reflects the reduced dielectric

screening of the Coulombic repulsion between paired electrons which is due to the protrusion of the orbital into a half space of SiO_2. A further contribution to this increase may come from the smaller lattice relaxation permitted by crystalline back bonds as compared to that determined by the matrix of lower average coordination in a-Si:H. In summary, the interfacial deep levels characteristic of thermally oxidized silicon correspond to the electronic transitions of the trivalent silicon defect at the Si-SiO_2 interface.

IV. CORRELATION EFFECTS IN DEEP-LEVEL MEASUREMENTS

The results presented in Sect. II and the model discussed in the last section introduce a new concept into the study of interface states, namely, the effect of electron-electron correlation on the measurement of interface-state distributions. Correlation energies have been determined for the bulk gap states in amorphous silicon[13,15] and amorphous germanium,[17] and correlation effects have been theoretically discussed with regard to field-effect measurements on amorphous semiconductors.[18]

In the present study a simple model was formulated and analyzed in order to examine the effect of a correlation energy on the measurement of interface-state distributions by charge transient spectroscopy. The results presented in Sect. II suggest that the P_b center possesses two transition levels in the silicon bandgap: the $0 \rightarrow 1$ electron transition at $E_V + 0.3$ eV and the $1 \rightarrow 2$ electron transition at $E_C - 0.25$ eV, with an effective correlation energy $U_0 \approx 0.6$ eV. In terms of hole occupancy, the total areal densities of P_b centers which are occupied with no holes, one hole, and two holes will be designated P_0, P_1, and P_2, respectively. Then the closure relation for defect occupancy is

$$N_S = P_0 + P_1 + P_2, \tag{1}$$

where N_S is the total areal density of P_b centers. The singly occupied center is charge neutral, and the unoccupied (P_0) and doubly occupied (P_2) centers are negatively and positively charged, respectively. The analysis of the transient response of an MOS capacitor with such amphoteric interface states follows that presented in Ref. 8 (see the appendix) for interface-state distributions arising from uncoupled defect levels. Here, the equations are derived for time-resolved transient-current and transient-voltage spectroscopies[8,19] and for CC-DLTS.[8]

The interface-state distribution in Fig. 1 suggests that each charge transition of the P_b centers can occur with a distribution of energies. It is physically plausible and consistent with these results to assume that the transitions are Gaussian distributions, with peaks at energies E_1 and E_2 above the silicon valence-band maximum which are separated by a fixed correlation energy U_0, i.e., $E_1 - E_2 \equiv U_0$. Then the two charge-transition distributions may be expressed as

$$N_i(E) = [N_S/(\sigma_G\sqrt{2\pi})] \exp[-(E-E_i)^2/(2\sigma_G{}^2)], \quad i = 1, 2, \qquad (2)$$

where σ_G is the standard deviation of the Gaussian distributions. The deep-level continuum analysis of Ref. 8 is applicable for $\sigma_G > kT$; the distributions in Fig. 1 indicate a $\sigma_G \approx 120$ meV. It should be noted that the Gaussian distributions only provide a focus for discussion; the results to be derived below do not depend on the specific functional form of the interface-state continuum.

At time $t = 0^+$ after a voltage pulse which saturates the interface states with holes (on a p-type MOS capacitor), all of the centers are doubly occupied with holes. The first of these paired holes is emitted with an emission

coefficient

$$e_{P2} = \sigma_{P2} v_P N_V e^{-E/kT}, \tag{3}$$

where σ_{P2} is the capture cross section for the 1→2 hole transition, v_P is the thermal velocity of free holes, N_V is the effective density of states in the valence band, k is Boltzmann's constant, and T is the absolute temperature. The emission coefficient for the second hole is

$$e_{P1} = \sigma_{P1} v_P N_V e^{-E/kT}, \tag{4}$$

where σ_{P1} is the capture cross section for the 0→1 hole transition. Spin degeneracy factors which also appear in the emission coefficients have been combined with the capture cross sections in this formulation. The first-order rate equations for hole emission from these coupled levels are

$$\partial p_2(E,t)/\partial t = -e_{P2}(E)\, p_2(E,t), \tag{5}$$

$$\partial p_1(E,t)/\partial t = e_{P2}(E - U_0)\, p_2(E - U_0, t) - e_{P1}(E)\, p_1(E,t), \tag{6}$$

$$\partial p_0(E,t)/\partial t = e_{P1}(E)\, p_1(E,t), \tag{7}$$

where p_2, p_1, and p_0 are the densities (per unit energy per unit area) of P_b centers containing two, one, and no holes, respectively. This formulation specifically assumes that there is no direct charge exchange between P_b centers, which appears justified given their mean separation of ~10 nm and the high degree of localization of the trapped charge on the deep-level defect. The initial conditions for the rate equations are $p_2(E,0) = N_2(E)$ and $p_1(E,0) = p_0(E,0) = 0$. The solutions to Eqs. (5) and (6) are

$$p_2(E,t) = N_2(E) \exp[-e_{P2}(E)\, t], \tag{8}$$

$$p_1(E,t) = R_0\, N_2(E - U_0) \{\exp[-e_{P1}(E)\, t] - \exp[-e_{P2}(E - U_0)\, t]\}, \tag{9}$$

where

$$R_0 \equiv e_{P2}(E - U_0)/[e_{P2}(E - U_0) - e_{P1}(E)]$$

$$= [1 - (\sigma_{P1}/\sigma_{P2}) \exp(-U_0/kT)]^{-1}. \quad (10)$$

It will be assumed that the capture cross sections are independent of energy so that R_0 is also a constant. The expression for $p_0(E,t)$ may be obtained from closure but is not needed for the analysis below.

The transient response of the capacitor depends on the change of charge in interface states with time. The net charge in interface states at time t is

$$Q_{it}(t) = q[P_2(t) - P_0(t)]. \quad (11)$$

In transient-current spectroscopy the current density at time t after a trap-filling voltage pulse, $j_{it}(t)$, depends on the interface charge as follows:[8]

$$j_{it}(t) = (1 - C_{HF}/C_{ox})\, dQ_{it}/dt, \quad (12)$$

where C_{HF} is the high-frequency capacitance for a given gate voltage V_G and C_{ox} is the oxide capacitance (in the following discussion, both capacitances are per unit area). Similarly, in time-resolved transient voltage spectroscopy, the derivative of the gate voltage varies with interface charge as follows:[8]

$$dV_G/dt = C_{ox}^{-1} dQ_{it}/dt \quad (13)$$

Finally, for CC-DLTS the expression is[8]

$$\Delta V_G = C_{ox}^{-1} [Q_{it}(t_1) - Q_{it}(t_2)], \quad (14)$$

where t_1 and t_2 are two specified delay times after the voltage pulse. From

Eq. (1) the time rate of change of interface charge appearing in Eqs. (12) and (13) is

$$dQ_{it}/dt = q(2dP_2/dt + dP_1/dt), \tag{15}$$

and similarly the charge difference factor in Eq. (14) may be expressed as

$$Q_{it}(t_1) - Q_{it}(t_2) \equiv \Delta Q_{it} = q\{2[P_2(t_1) - P_2(t_2)] + [P_1(t_1) - P_1(t_2)]. \tag{16}$$

To complete the formulation for the several charge–transient spectroscopies, the distributions of occupied centers $p_i(E,t)$ are formally related to the integrated densities $P_i(t)$ as follows:

$$P_i(t) \equiv \int_{E_V}^{E_C} p_i(E,t)dE, \; i = 1, 2. \tag{17}$$

The above formulation thus specifies the time evolution of charge in amphoteric interface states after complete interface trap filling.

In terms of the charge–transition distributions $N_i(E)$, Eqs. (15) and (16) yield,

$$dQ_{it}/dt = q[-(2-R_0)N_2(E_{m2}^*) - R_0N_1(E_{m1}^*)]\,kT/t, \tag{18}$$

and

$$\Delta Q_{it} = q[(2-R_0)N_2(E_{m2}) + R_0N_1(E_{m1})]\,kT\,ln(t_2/t_1), \tag{19}$$

respectively, where

$$E_{mi}^* \equiv kT\,ln(\sigma_{Pi}v_PN_V\,t/\beta_m^*), \tag{20}$$

and

$$E_{mi} \equiv kT\,ln(\sigma_{Pi}v_PN_V\,t_1/\beta_m), \tag{21}$$

with $i = 1, 2$. Numerical values of β_m^* and β_m are given in Ref. 8.

The uniqueness of amphoteric centers as compared to uncoupled levels may be illustrated by considering the time dependences of the integrated densities of the two charge-transition levels. For uncoupled distributions both P_2 and P_1 decrease monotonically with time, under emission dominated kinetics. However, for the amphoteric centers the initial conditions are $P_2(0) = N_S$ and $P_1(0) = P_0(0) = 0$, and as $t \rightarrow \infty$ the densities are $P_2 \rightarrow 0$, $P_1 \rightarrow 0$, and $P_0 \rightarrow N_S$. Thus the integrated density of singly occupied P_b centers, $P_1(t)$, passes through a maximum at some non-zero finite time $t_{P1,max}$ which depends on the distribution parameters as follows:

$$t_{P1,max} = [\beta_m^*/(v_P N_V \sqrt{\sigma_{P1}\sigma_{P2}})] \exp[(E_1 + E_2)/(2kT)]. \quad (22)$$

The interface trap emission energies at this time are

$$E_{mi}^*(t_{P1,max}) = (kT/2)\, ln(\sigma_{Pi}/\sigma_{Pj}) + (E_1 + E_2)/2, \quad (23)$$

with i, j = 1, 2 and $i \neq j$.

An experiment which directly detects P_1 could demonstrate the amphoteric nature of the center through its unique time dependence. For example, the ESR signal due to unpaired electrons in P_b centers is a direct measure of P_1. If the spin signal could be time resolved with adequate signal to noise, it would directly yield the time dependence of P_1 for comparison with the model prediction. Such an experiment appears to present practical difficulties but cannot be ruled out. However, the parameters for the P_b center suggested in Sects. II and III, particularly the large effective correlation energy, predict that either the time regime or the temperature would have to be changed to time resolve each of the two transitions, as further discussed below.

As evident from Eqs. (12) - (14), (18), and (19), the existence of amphoteric defects can alter the analysis of DLTS measurements of interface

states from that based on uncoupled levels. However, the parameters suggested for the P_b center in Sects. II and III permit a revealing simplification. The large effective correlation energy ($U_0 \simeq 0.6$ eV) and the low temperatures of the DLTS measurement ($\lesssim 300$ K) yield $U_0 >>> kT$ so that $R_0 \rightarrow 1$ and Eq. (18) becomes

$$dQ_{it}/dt \simeq -q[N_2(E_{m2}^*) + N_1(E_{m1}^*)]\, kT/t. \tag{24}$$

A similar expression follows from Eq. (19) for ΔQ_{it}. In addition, due to the small amount of overlap of the two transition distributions (i.e., for $U_0 >> \sigma_G$) it follows that in the temperature–time regime in which the first transition appears in the range of the rate window of the spectrometer (e.g., $E_{m2}^* \approx 0.3$ eV for the $2 \rightarrow 1$ hole transition in p-type MOS devices) the relative densities of the levels are $N_2(E_{m2}^*) >> N_1(E_{m1}^*)$ and Eq. (24) simplifies to

$$dQ_{it}/dt \simeq -qN_2(E_{m2}^*)\, kT/t. \tag{25}$$

Only near midgap are the densities comparable so that Eq. (24) becomes a more accurate description of the transient response; however, near midgap other complications arise as discussed in Ref. 8.

Equations (24) and (25) are identical to the expressions which would rigorously apply if the two peaks in the interface-state distribution (Fig. 1) arose from uncoupled levels. Thus it is demonstrated that due to the large effective correlation energy of the P_b center and the close proximity of each transition level to a band edge, the DLTS signal is dominated by only a single charge transition (i.e., the $E_V + 0.3$ eV peak in p-type MOS capacitors and the $E_C - 0.25$ eV peak in n-type devices) so that Eq. (25), or equivalently the analysis of Ref. 8 which assumes uncoupled levels, is an accurate approximation and produces a self-consistent interface-state distribution (Fig. 1).

ACKNOWLEDGMENT

Co-investigators of the work presented in this report are D. K. Biegelsen and M. D. Moyer. S. T. Chang (Princeton University) also contributed to sample preparation and measurements. The research was conducted in collaboration with E. H. Poindexter and P. J. Caplan of the U. S. Army Electronics Technology and Devices Laboratory. We are pleased to acknowledge helpful discussions with J. D. Chadi and R. A. Street during the course of this work. Special thanks are also extended to V. Moffat for typing and editing this report.

REFERENCES

1. N. M. Johnson, D. J. Bartelink, and M. Schulz, *The Physics of SiO_2 and its Interfaces* (Pergamon, New York, 1978), ed. S. T. Pantelides, pp. 421 – 427.

2. N. M. Johnson, D. J. Bartelink, and J. P. McVittie, J. Vac. Sci. Technol. 16, 1407 (1979).

3. N. M. Johnson, D. K. Biegelsen, and M. D. Moyer, *The Physics of MOS Insulators* (Pergamon, New York, 1980), eds. G. Lucovsky, S. T. Pantelides, and F. L. Galeener, pp. 311 – 315.

4. P. J. Caplan, E. H. Poindexter, B. E. Deal, and R. R. Razouk, J. Appl. Phys. 50, 5847 (1979).

5. R. B. Laughlin, J. D. Joannopoulos, and D. J. Chadi, Phys. Rev. B 21, 5733 (1980).

6. E. H. Poindexter, P. J. Caplan, J. J. Finnegan, N. M. Johnson, D. K. Biegelsen, and M. D. Moyer, *The Physics of MOS Insulators* (Pergamon, New York, 1980), eds. G. Lucovsky, S. T. Pantelides, and F. L. Galeener, pp. 326 – 330.

7. C. Brunstrom and C. Svensson, Solid State Commun. 37, 399 (1981).

8. N. M. Johnson, J. Vac. Sci. Technol. 21, 303 (1982).

9. N. M. Johnson, Appl. Phys. Lett. 34, 802 (1979).

10. E. H. Nicollian and J. R. Brews, *MOS Physics and Technology* (Wiley, New York, 1982).

11. D. K. Biegelsen, *Nuclear and Electron Resonance Spectroscopies Applied to Materials Science* (Elsevier, New York, 1981), eds. E. N. Kaufmann and G. K. Shenoy, pp. 85 - 94.

12. R. A. Street and D. K. Biegelsen, Solid State Commun. 33, 1159 (1980).

13. H. Dersch, J. Stuke, and J. Beichler, Phys. Stat. Solid. (b) 105, 265 (1981).

14. D. K. Biegelsen, *Defects in Semiconductors* (Elsevier, New York, 1983), eds. S. Mahajan and J. W. Corbett, in press.

15. W. B. Jackson, Solid State Commun. 44, 477 (1982).

16. The P_b spin density was found to be unchanged (within $\pm$ 5%) in the investigated temperature range of 30 to 300 K. This fact, coupled with the pinning of the Fermi energy between the two defect levels (in the absence of an applied voltage), implies a *lower bound* for the effective correlation energy of $\geq$ 50 meV.

17. M. Stutzmann, J. Stüke, and H. Dersch, Phys. Stat. Sol. (b) 115, 141 (1983).

18. L. Schweitzer, M. Grünewald, and H. Dersch, Solid State Commun. 39, 355 (1981).

19. N. M. Johnson, Appl. Phys. Lett. 42, 981 (1983).

FIGURE CAPTIONS

Fig. 1. Interface-state distribution measured by transient current spectroscopy on thermally oxidized, unannealed, (111) - oriented silicon. The distribution is a composite of measurements on identically processed n - type and p - type MOS capacitors. The energies E_V and E_C identify the silicon valence - band maximum and conduction-band minimum, respectively.

Fig. 2. Comparison of isochronal anneals of the P_b spin density and the density of the $E_V + 0.3$ eV peak in the interface-state distribution on (111)-oriented, n-type silicon. The specimens were coated with aluminum and annealed in vacuum for 15 min. The circles are the characteristic peak densities, and the squares are the spin densities. Both densities are normalized to the maximum (unannealed) values.

Fig. 3. Dependence of the P_b spin density, S, normalized to the maximum density S_{max}, on applied gate voltage V_G in a large-area MOS capacitor. The data are a compilation of measurements from four identically prepared capacitors, and each datum is the peak - to - peak amplitude of the P_b (derivative) absorption spectrum averaged over ten or more scans and corrected for any background signal.

Fig. 4. Schematic energy-band diagram for the density of gap states in hydrogenated amorphous silicon.

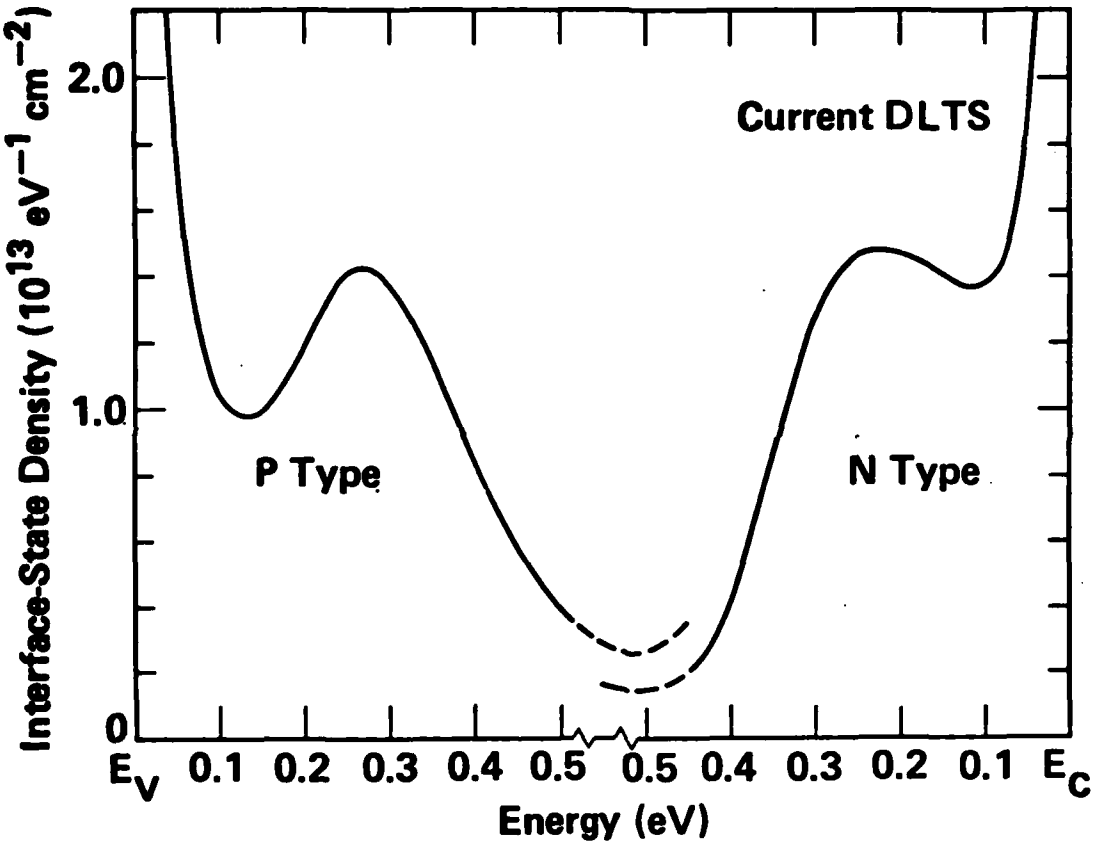

Interface-State Density (10^{13} eV^{-1} cm^{-2})
2.0
1.0
0
Current DLTS
P Type
N Type
E_V 0.1 0.2 0.3 0.4 0.5 0.5 0.4 0.3 0.2 0.1 E_C
Energy (eV)

Fig. 1.

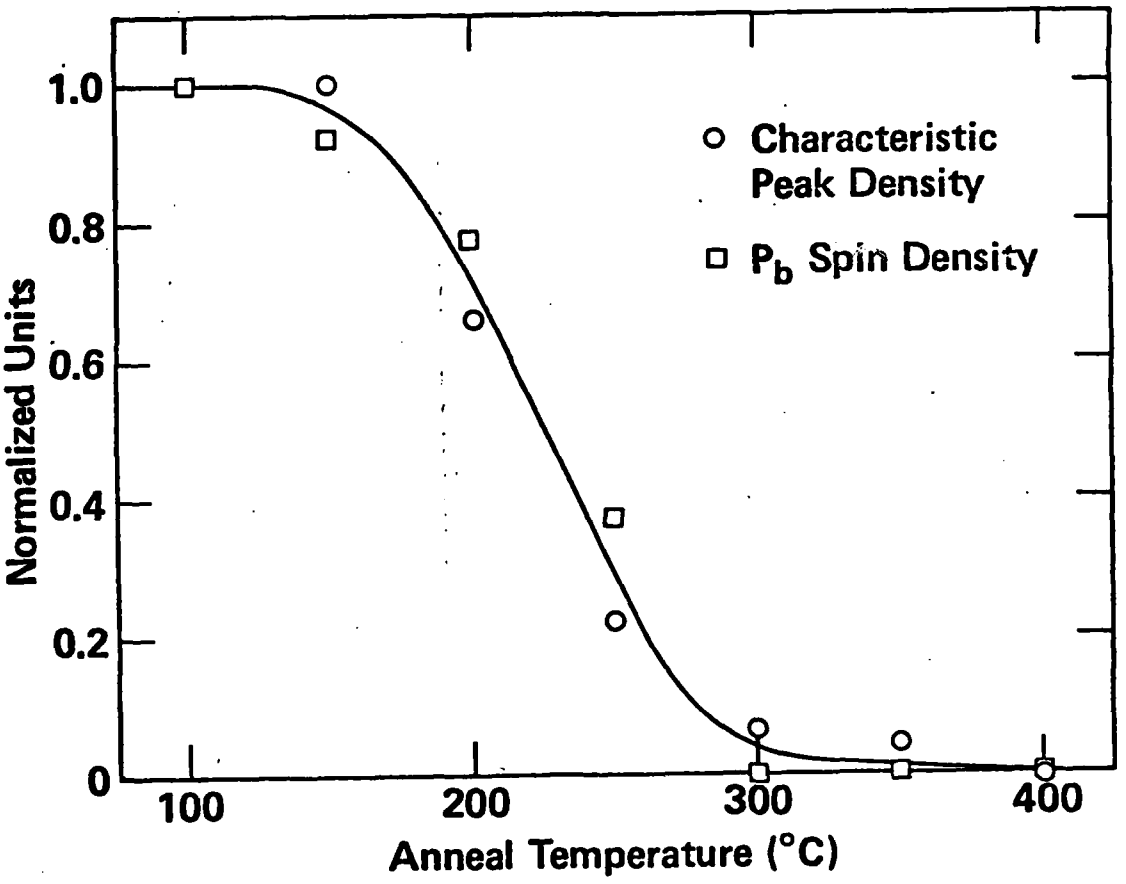

1.0
0.8
0.6
0.4
0.2
0
Normalized Units
100
200
300
400
Anneal Temperature (°C)
Characteristic Peak Density
P_b Spin Density

Fig. 2.

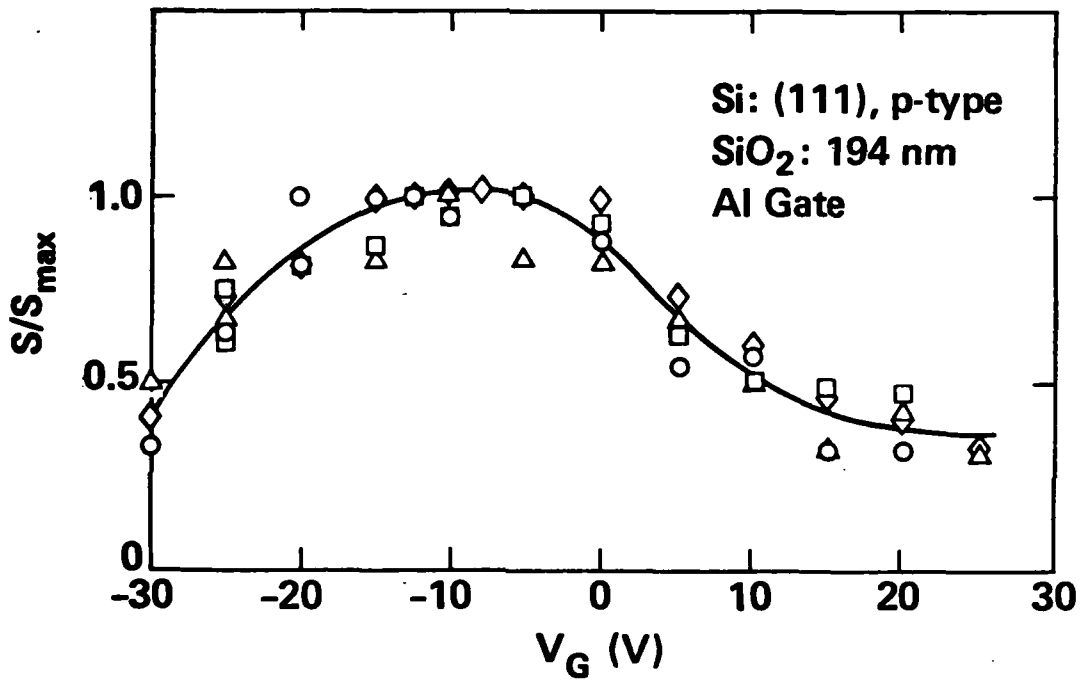

Si: (111), p-type
SiO_2: 194 nm
Al Gate
S/S_{max}
1.0
0.5
0
-30
-20
-10
0
10
20
30
V_G (V)

Fig. 3.

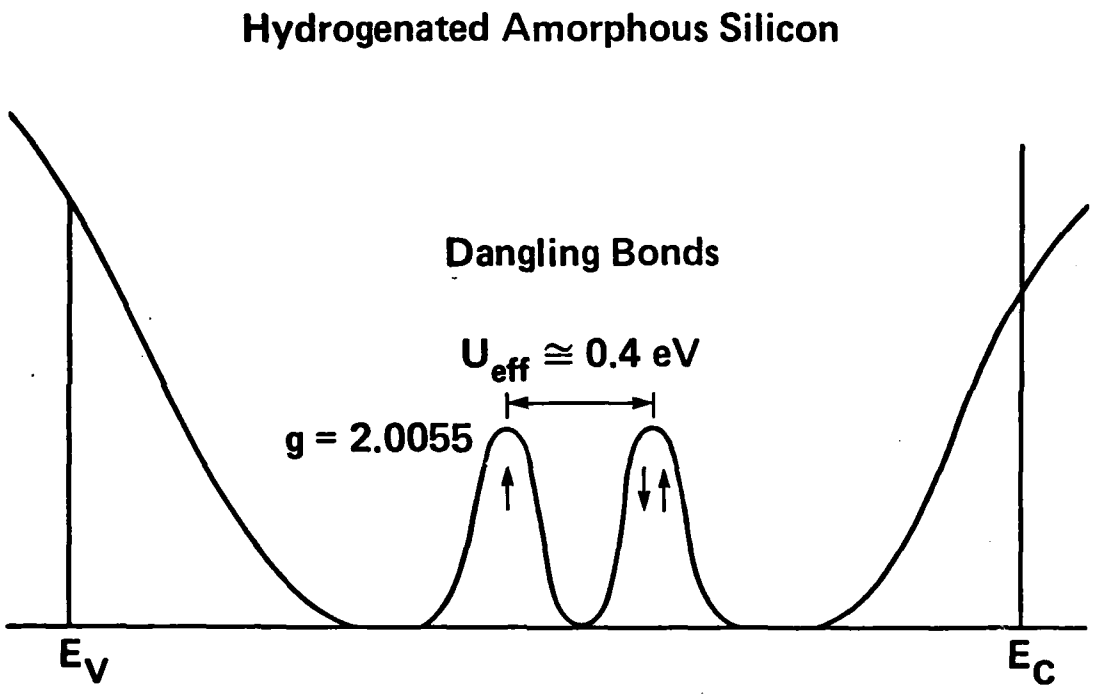

Hydrogenated Amorphous Silicon
Dangling Bonds
$U_{eff} \cong 0.4$ eV
g = 2.0055
E_V
E_C

Fig. 4.

APPENDIX:

MEASUREMENT OF SEMICONDUCTOR-INSULATOR INTERFACE STATES BY CONSTANT-CAPACITANCE, DEEP-LEVEL TRANSIENT SPECTROSCOPY

Measurement of semiconductor–insulator interface states by constant-capacitance, deep-level transient spectroscopy

N. M. Johnson

Xerox Palo Alto Research Center, Palo Alto, California 94304

(Received 27 October 1981; accepted 14 April 1982)

Localized electronic states at the semiconductor–insulator interface adversely affect the operation of insulated-gate, field-effect devices. Characterization of interface states provides essential information for minimizing their effect through process optimization, for predicting device performance, and ultimately for microscopic identification of interface defects. This paper reviews the application of deep-level transient spectroscopy (DLTS) for characterizing interface states on metal–insulator–semiconductor capacitors, with emphasis on the constant-capacitance (CC) mode of measurement. The DLTS measurement yields both the energy distribution of interface states and their cross section for capturing free carriers. In addition, it has the versatility of being applicable to both interface and bulk defect characterization. The CC-DLTS technique offers the combined features of high sensitivity ($< 1 \times 10^9$ eV^{-1} cm^{-2}), minimum signal distortion at high defect densities, high energy resolution, and the determination of dynamic properties. After a description of the measurement system and experimental procedures, the theoretical basis is developed for data reduction of majority-carrier-dominated transients for the following cases: (1) under saturating-pulse conditions and (2) with Fermi-level controlled trap occupancy. Under the first topic is included a summary of transient-current spectroscopy, and the second is illustrated with the energy-resolved DLTS technique. The presentation includes an analysis of the effect of surface generation on the DLTS measurement of interface states near the semiconductor midgap and an analysis of the limits of applicability of the *transient-capacitance* mode for DLTS measurement of interface states. The techniques are illustrated with measurements of electronic defect levels at the Si–SiO_2 interface.

PACS numbers: 73.40.Qv, 71.55. – i

I. INTRODUCTION

Localized electronic states at the semiconductor–insulator interface adversely affect the operation of insulated-gate, field-effect devices. Measurement of interface states provides essential information for minimizing their effect through process optimization and for predicting device performance; and combining electrical characterization with other physical measurements (e.g., electron spin resonance) may also permit microscopic identification of electronic defects at interfaces. This paper describes the application of deep-level transient spectroscopy (DLTS) for characterizing interface states, with emphasis on the constant-capacitance (CC) mode of measurement. For this application, constant-capacitance DLTS features high sensitivity for interface-defect detection, minimum signal distortion at high defect densities, high energy resolution, and the determination of dynamic properties.[1,2] In addition, it is a versatile experimental technique being applicable to both interface and bulk defect characterization.

The basic materials configuration for insulated-gate, field-effect devices and interface evaluation is the metal–insulator–semiconductor (MIS) structure; when the insulator is an oxide, the structure is commonly designated MOS. The physical principles and operation of MIS devices have been thoroughly discussed in the literature.[3–5] Similarly, measurements of interface states with small-signal capacitance and conductance techniques (in the frequency domain) have been comprehensively reviewed[5,6] and are not discussed here.

Deep-level transient spectroscopy is a large-signal time-domain technique developed by Lang[7] to characterize bulk deep levels in crystalline semiconductors. In its conventional form the time-varying capacitance of a *p-n* junction or Schottky-barrier diode is used to detect capture and emission of charge carriers at bulk deep levels.[8] The MIS structure has also been used for bulk-defect characterization.[9]

Interface states constitute a unique form of defect in crystalline semiconductors. In bulk single-crystal semiconductors, deep levels appear at discrete energies and are spatially distributed. At the interface between a crystalline semiconductor and an amorphous insulator, deep levels are continuously distributed in energy and spatially localized. The application of conventional transient-capacitance DLTS for interface-state measurements, on field-effect transistors and MIS capacitors, has recently been reviewed.[10] In addition, current transients have been used in deep-level spectroscopy for interface-state characterization.[11,12] In this paper features of both techniques are discussed in relation to CC–DLTS.

The qualitative features of the DLTS measurement of interface states are illustrated in Fig. 1 with energy-band dia-

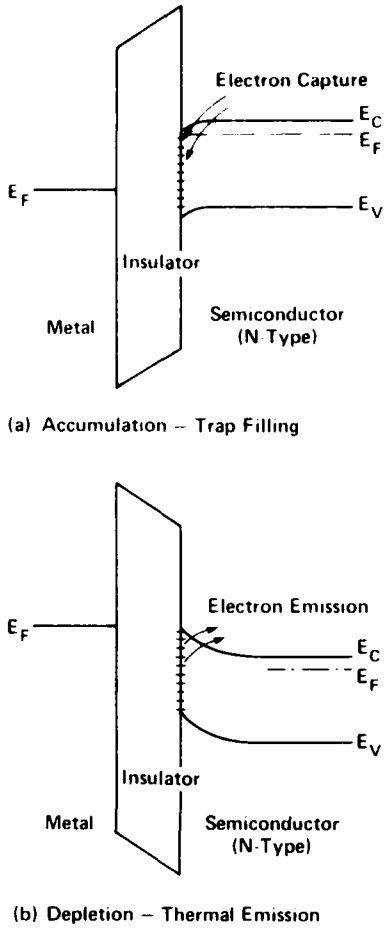


FIG. 1. Energy-band diagrams for an MIS device on an *n*-type semiconductor for (a) pulsed accumulation bias and (b) nonequilibrium depletion bias.

grams for an MIS structure on an *n*-type semiconductor. The experimental procedure involves the application of a gate voltage which consists of the superposition of a dc value, used to bias the device in depletion, and periodic voltage pulses. The gate voltage is pulsed to drive the device into accumulation in order to populate interface states with electrons (majority carriers), as shown in Fig. 1(a). After a trap-filling pulse with the gate voltage returned to the depletion value, the interface states possess a nonequilibrium distribution of trapped charge which relaxes by thermal emission of electrons to the conduction band, as depicted in Fig. 1(b). The DLTS measurement consists of monitoring the interface-trap emission rate over a range of temperatures. From the analysis can be deduced the interface-state distribution in the semiconductor bandgap and the cross section of interface states for capturing free carriers.

This paper reviews the experimental and analytical procedures for the application of constant-capacitance DLTS for interface-state measurement. The discussion focuses on the two-terminal MIS capacitor because it requires minimum fabrication and therefore offers maximum flexibility for interface materials studies. After a description of the measurement system and procedures, a unified analysis of the key experimental methods is presented with more detail than previously reported. The presentation includes an analysis of the effect of surface generation on the DLTS measurement of interface states near the middle of the semiconductor forbidden energy band. Also presented is an analysis of the limits of applicability of the transient-capacitance mode for interface-state measurement. In Sec. II are described the experimental techniques for CC–DLTS measurements. The theoretical basis for data reduction is developed in Sec. III with analyses of (1) majority-carrier emission after complete interface-trap filling, (2) emission after partial trap filling, and (3) the effect of minority-carrier emission and capture processes on the measurement of interface states near the semiconductor midgap. The first topic includes a discussion of transient-current spectroscopy for comparison with CC–DLTS. Under the second is discussed the use of the Fermi level to control trap occupancy, which is illustrated with the energy-resolved DLTS technique. In the third topic is presented an analysis of the effect of surface generation on the DLTS measurement of interface states near midgap. The techniques and analyses are illustrated with measurements of electronic defect levels at the Si–SiO_2 interface. In the appendix an analysis of transient-capacitance DLTS is presented and used to evaluate the limits of applicability of this mode for interface-state measurement.

II. EXPERIMENTAL TECHNIQUES

A block diagram of the apparatus for constant-capacitance DLTS is shown in Fig. 2. The measurement system is based on the transient-capacitance system proposed by Lang[7] and consists of a capacitance bridge with fast transient response, a pulse generator for rapidly changing sample bias, a dual-gated signal integrator for monitoring and signal

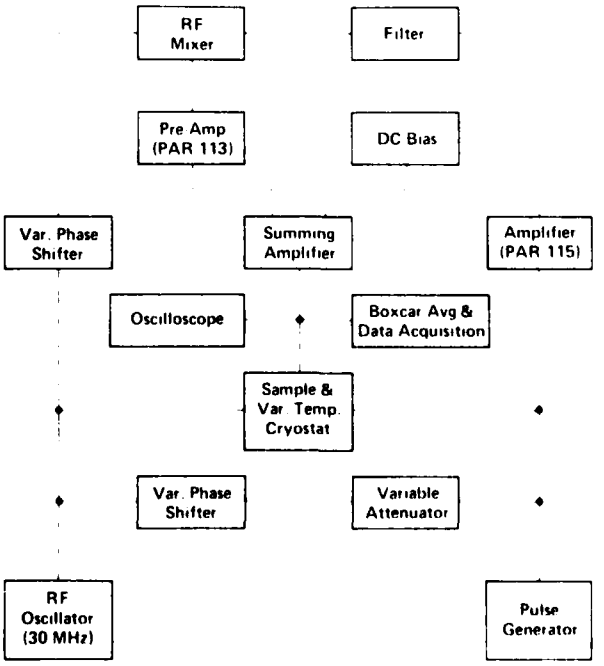


FIG. 2. Block diagram of apparatus for constant-capacitance, deep-level transient spectroscopy.

averaging the transient response, and a variable-temperature cryostat in which the device is mounted. For the constant-capacitance mode of measurement, the basic system has been expanded to include feedback circuitry which maintains the capacitance of an MIS device at a constant value by dynamically varying the gate voltage. In the transient-capacitance mode, the bridge is balanced for a given dc depletion bias, and the capacitance transient is monitored at the output of the preamplifier (PAR 113). For constant-capacitance measurements, this signal is sent to a summing amplifier to provide an applied voltage consisting of the algebraic sum of the dc depletion bias and the time-varying voltage required to maintain the capacitance at the depletion value. In this form of operation, the branch of the bridge in parallel with the test device, which consists of the series-connected, variable-phase shifter and variable attenuator, in effect provides the reference signal for the feedback circuit. Successful operation of the feedback system requires proper choice of signal polarity for negative feedback and adjustment of gain and bandwidth for stable operation. A computer-controlled automated system for transient-capacitance and constant-capacitance DLTS has been described,[13] and the constant-capacitance mode of measurement can be readily implemented on DLTS systems utilizing commercial capacitance meters.

The voltage transient response is schematically illustrated in Fig. 3 with a hf capacitance–voltage characteristic for an MIS capacitor on an n-type semiconductor. The device is initially biased in depletion to establish a depletion capacitance C_{hf} at gate bias V_{dc}. A voltage pulse V_{pulse} drives the capacitor into accumulation during which interface states are populated with electrons. After the pulse the depletion

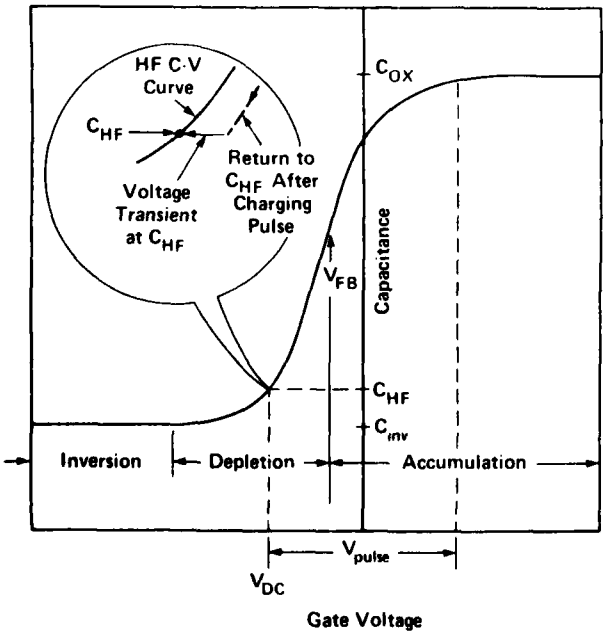


FIG. 3. High-frequency capacitance–voltage characteristic for an MIS capacitor on an n-type semiconductor. The insert schematically illustrates the relaxation of the gate voltage toward the depletion bias V_{dc} after a voltage pulse V_{pulse} into accumulation during a CC-DLTS measurement.

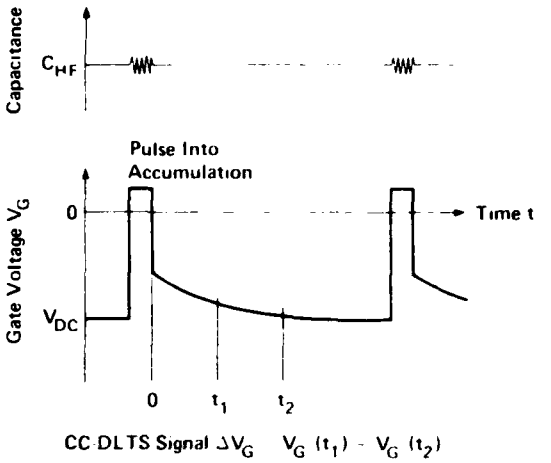


FIG. 4. Schematic diagrams of the capacitance and gate-voltage waveforms in a CC-DLTS measurements on MIS capacitors

bias is servoed to maintain C_{hf} constant as the interface-state population returns to the equilibrium distribution corresponding to V_{dc}.

The capacitance and voltage waveforms are schematically represented in Fig. 4. After a pulse into accumulation, the capacitance is held constant by dynamically varying the gate voltage as the charge in interface states relaxes to equilibrium. The CC-DLTS signal ΔV_G is obtained by forming the difference of the gate voltages measured at two delay times t_1 and t_2 after a charging pulse. With the system in Fig. 2 this is readily achieved by periodically pulsing the gate voltage and monitoring the transient response after each pulse with the boxcar averager, which is equipped for dual-gated signal integration and hence also provides good noise discrimination by time averaging. The signal is recorded in a temperature scan to form a trap-emission spectrum.

III. ANALYSIS

Charge relaxation in interface states, after a gate-voltage pulse from depletion to accumulation, contributes to the transient response of an MIS device which can be experimentally observed with the techniques described in the previous section. In the analysis presented here it is assumed that emission and capture processes at the interface dominate the transient response. Data analysis yields the distribution in energy of the interface states and their cross section for capturing free carriers. The semiconductor–insulator interface is considered to be essentially abrupt and to possess a continuous distribution of electronic energy levels in the semiconductor bandgap. The defect density is assumed to be sufficiently low (i.e., $\leq 10^{13}$ cm $^{-2}$) that the Fermi level is not pinned at the interface. There is no charge injection and trapping in the insulator, and there are no deep levels in the bulk of the semiconductor; this latter assumption is discussed in Sec. IV. The conditions of the model are, in general, well satisfied in MIS devices consisting of a silicon-dioxide layer on single-crystal silicon. In addition, lateral inhomogeneities, which give rise to fluctuations in the semiconductor surface potential, can be ignored for majority-carrier emis-

sion after complete trapping filling; this is further discussed in Sec. IV.

A. CC–DLTS signal

The CC–DLTS signal is linearly dependent on the difference in net charge in interface states at the two delay times t_1 and t_2 after a charging pulse. From Fig. 4 the CC–DLTS signal is formed as follows:

$$\Delta V_G \equiv V_G(t_1) - V_G(t_2). \tag{1}$$

The linear dependence on interface charge is readily shown by examining the high-frequency capacitance of an MIS device. For a depletion bias, the high-frequency device capacitance C_{hf} is the series combination of the insulator capacitance C_i and the semiconductor depletion capacitance C_D, with C_{hf} held constant during the trap-emission transient response

$$C_{hf} = C_i C_D/(C_i + C_D) = \text{constant}. \tag{2}$$

Since the insulator capacitance is independent of bias, it follows that the depletion capacitance remains constant during the transient response. The applied gate bias may be expressed as an effective voltage V'_G consisting of the sum of the potential difference across the insulator V_i and the semiconductor surface potential ψ_s

$$V'_G = V_i + \psi_s. \tag{3}$$

This expression specifically neglects the metal–semiconductor work function difference and space charge in the insulator. From Gauss' law, V_i depends on the net charge in interface states Q_{is} and charge in the semiconductor depletion layer Q_s as follows:

$$V_i = -A(Q_s + Q_{is})/C_i, \tag{4}$$

where A is the gate area and the charge densities are per unit area. With the depletion capacitance held constant during the transient response, the semiconductor charge density and surface potential also remain constant. Then, from Eq. (3) the change in gate voltage required for constant C_{hf} appears only across the insulator. Hence, from Eqs. (1), (3), and (4)

$$\Delta V_G = (A/C_i)[Q_{is}(t_1) - Q_{is}(t_2)]. \tag{5}$$

Thus, the linear dependence obtains for all interface trap densities, and the proportionality factor in Eq. (5) is simply the inverse of the insulator capacitance per unit area. It is therefore a fixed parameter for a given insulating layer. In contrast, with *transient-capacitance* measurements of interface states on MIS devices, the proportionality factor is a function of both materials and device parameters and, in general, varies with temperature. Since the DLTS emission spectrum is recorded in a temperature scan, this can introduce significant distortion. In the appendix this proportionality factor is derived, and the limits of applicability of the transient-capacitance mode for interface-state measurement is evaluated.

B. Complete trap filling

The extreme nonequilibrium condition for monitoring the transient response of an MIS capacitor is achieved by totally populating the interface states with majority carriers by the end of the voltage pulse. With the capacitor in thermal equilibrium and biased in depletion, the Fermi level at the interface intersects the semiconductor bandgap near midgap. When the device is pulsed into strong accumulation, the interface states are readily populated with majority carriers essentially to the majority carrier band edge. For electrons in an n-type semiconductor, the capture coefficient is

$$c_n = \sigma_n v_n n_s, \tag{6}$$

where σ_n is the capture cross section for electrons, v_n is the mean thermal velocity for electrons, and n_s is the free carrier (electron) concentration at the semiconductor–insulator interface. In general, n_s decreases with time during the charging pulse,[14] however this effect is not significant for complete trap filling. In strong accumulation n_s approaches degenerate concentrations which correspond to the Fermi level intersecting the interface near the conduction band minimum. For a silicon MOS capacitor pulsed into strong accumulation, such that $n_s \sim 10^{19}\ \text{cm}^{-3}$, and interface states with $\sigma_n = 10^{-15}\ \text{cm}^2$, the capture time $(1/c_n)$ is of the order of 10^{-10} s. Thus, complete trap filling, which produces a saturated CC-DLTS signal at all temperatures, is readily achieved with commonly used pulse widths of a microsecond or more.

C. Majority-carrier emission

After the voltage pulse, electrons are emitted from interface states to the conduction band as the charge distribution relaxes toward the equilibrium distribution corresponding to the given depletion bias. With $n_{is}(E,t)$ defined as the density of electrons trapped in interface states at energy E below the semiconductor conduction-band minimum and at time t after the voltage pulse, the rate of emission is

$$dn_{is}/dt = -e_n n_{is}(E,t), \tag{7}$$

where e_n is the electron emission coefficient which from considerations of detailed balance may be expressed as

$$e_n = (\sigma_n v_n N_c/g_{is})\exp[-E/(kT)]. \tag{8}$$

Here, N_c is the effective density of states in the semiconductor conduction band, k is Boltzmann's constant, and T is the absolute temperature. Without detailed knowledge of the interface states in a given material, it is generally necessary to assume a value of unity for the spin degeneracy factor g_{is}. With the interface states completely occupied with electrons at $t = 0$, Eq. (7) yields $n_{is} = N_{is}(E)\exp(-e_n t)$, where $N_{is}(E)$ is the interface-state density at energy E. The transient response of an MIS capacitor during majority-carrier emission is evaluated for constant-capacitance conditions in the following subsection (III C 1) and for transient-current spectroscopy in Sec. III C 2. The transient-capacitance mode of measurement is analyzed in the appendix.

1. *CC–DLTS*

From Eq. (5), the dependence of the CC–DLTS signal on electron emission from a continuous distribution of interface states is given by

$$\Delta V_G = (A/C_i)\int qN_{is}(E)[\exp(-e_n t_1) - \exp(-e_n t_2)]dE, \tag{9}$$

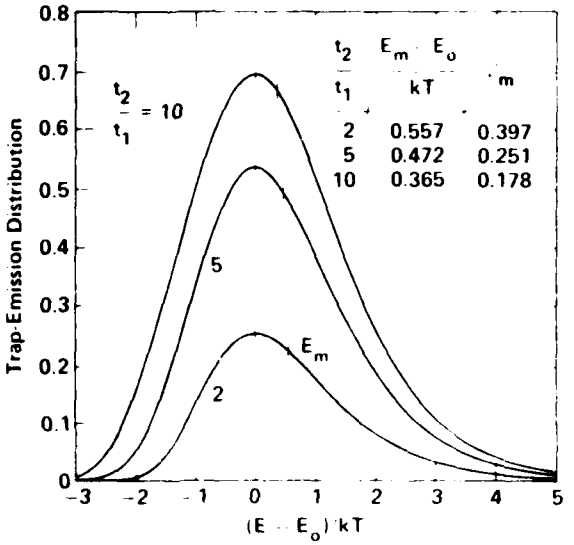


FIG. 5. Trap-emission distributions for commonly used ratios of the DLTS delay times. The mean energy E_m is marked on each distribution and listed in the insert with the proportionality constant β_m for the mean emission rate window, $\beta_m = (e_m t_1)|_{t_2/t_1}$.

where q is electronic charge and the integration is over the interval of the semiconductor bandgap in which majority-carrier emission dominates the relaxation process.

To obtain an expression for ΔV_G which is useful for data reduction requires examination of the integrand in Eq.(9) and an additional assumption. The exponential terms in the bracketed portion of the integrand form a function, termed here the trap-emission distribution, which is shown in Fig. 5 for commonly used ratios of the delay times. The distribution is peaked at an emission rate e_0 given by

$$e_0 = (t_2 - t_1)^{-1} \ln(t_2/t_1), \tag{10}$$

which from Eq. (8) corresponds to an energy

$$E_0 = kT \ln(\sigma_n v_n N_C / e_0). \tag{11}$$

For a capture cross section which is independent of energy, the trap-emission distribution plotted versus the dimensionless variable $(E - E_0)/kT$ as in Fig. 5 depends only on the ratio t_2/t_1. This may be shown analytically with Eqs. (8), (10), and (11).[5] The distribution has a full width at half maximum of $\sim 2.5kT$. If the interface-state density varies slowly over this interval, it may be treated as a constant and taken outside the integral to yield an effective energy interval ΔE_n defined as follows:

$$\Delta E_n = \int_{\text{}} [\exp(-e_n t_1) - \exp(-e_n t_2)]\, dE. \tag{12}$$

If the interface-state density varies linearly with energy over the interval ΔE_n, then the density is evaluated at the mean energy E_m of the trap-emission distribution[15]; this is to be contrasted with the case of discrete energy levels in which trap parameters are determined at the peak energy E_0.[7] In Fig. 5 the mean energy is marked on each trap-emission distribution and specified in the insert in units of kT relative to E_0. Also listed are values for the proportionality constant β_m relating the mean emission rate coefficient e_m to the delay time t_1 for fixed values of t_2/t_1; that is, $\beta_m \equiv (e_m t_1)|_{t_2/t_1}$. These values apply for energy-independent capture cross sections. The difference between E_0 and E_m is less than kT and is generally neglected being less than the energy resolution of the measurement. Both e_0 and e_m depend only on the specified delay times of the spectrometer; either can be used to specify an emission rate window.

With the assumption that N_{ss} varies linearly with energy over the interval ΔE_n, Eq. (9) becomes

$$\Delta V_G = (A/C_i) q N_{ss}(E_m) \Delta E_n. \tag{13}$$

If the capture cross section is independent of energy, the effective energy interval is

$$\Delta E_n = kT \ln(t_2/t_1). \tag{14}$$

Thus, the CC–DLTS signal arises from the emission of majority carriers from interface states in an energy interval ΔE_n which is situated at energy E_m from the majority-carrier band edge. For a constant capture cross section, both E_m and ΔE_n increase linearly with temperature. For given delay times the energy resolution is therefore greatest at low temperatures where the sampled energy interval is closest to the conduction band and the interface-state density generally varies most rapidly (e.g., band-tail states). The larger energy interval at high temperatures provides enhanced sensitivity for detecting low densities of interface states which are often found near midgap. Densities $< 1 \times 10^9$ eV^{-1} cm^{-2} are readily measurable (e.g., see Sec. IV and Ref. 1).

A useful extension of the above analysis is to the case where the capture cross section varies with energy. For example, there is evidence that the cross section for electron capture at the Si–SiO_2 interface increases exponentially with trap depth near the silicon conduction band to a constant value over the central region of the silicon bandgap.[16,17] The effect of such an energy dependence on DLTS emission spectra has been considered in computer simulations[14] and can be evaluated analytically. For a capture cross section which varies as $\sigma_n(E) = \sigma_0 \exp(E/E_a)$, the effective energy interval is

$$\Delta E_n(\sigma_n) = [kT \ln(t_2/t_1)]/(1 - kT/E_a), \tag{15}$$

where σ_0 and E_a are constants. More generally, for an energy-dependent capture cross section the effective energy interval is given by[15]

$$\Delta E_n(\sigma_n) = [kT \ln(t_2/t_1)]/[1 - kT\,[d \ln \sigma_n(E)/dE]_{E = E_m}]. \tag{16}$$

The expression for ΔV_G in Eq. (13) remains applicable if the interface-state density can be assumed to vary linearly over the broader energy interval given by Eq. (15) as compared to Eq.(14). Thus, from Eqs. (14) and (15) as σ_n varies from an exponential dependence on energy near the silicon conduction band to a constant over the central region of the bandgap, the effective energy interval for interface-state detection increases first superlinearly and then with a linear dependence on temperature. The effect can be to introduce an emission peak at low temperatures in a DLTS spectrum, which, with the assumption of a constant capture cross section, would suggest a peak in the interface-state distribution near the conduction band even though the interface-state density is actually smoothly varying.[14] Measurement of en-

ergy-dependent cross sections is discussed in Sec. III D.

From Eq. (14), the requirement that N_{is} vary linearly over the interval ΔE_n at E_m dictates that the ratio of the delay times be kept small. On the other hand, the DLTS signal-to-noise ratio improves with increasing values of t_2/t_1. A practical trade off between sensitivity and energy resolution is to set $t_2 = 2t_1$. Application of the above analysis is illustrated in Sec. III D.

2. *Transient-current spectroscopy*

The key analytical results for transient-current spectroscopy can be summarized in the nomenclature introduced above for comparison with CC-DLTS. For electron emission from interface states with a fixed depletion bias, the current density at time t after a charging pulse is[11]

$$J(t) = [(C_i - C_{hf})/C_i] \int qN_{is}(E)e_n \exp(-e_n t)dE. \quad (17)$$

The function $e_n \exp(-e_n t)$ is sharply peaked at the emission rate $e'_0 = 1/t$. If the interface-state density varies linearly over the width of this peak the current transient is

$$J(t) = [(C_i - C_{hf})/C_i]qN_{is}(E'_m)\Delta E'_n/t, \quad (18)$$

where the effective energy interval $\Delta E'_n = kT$ for an energy-independent capture cross section. The mean energy E'_m may be obtained from $(E'_m - E'_0)/(kT) = 0.577$, and the proportionality constant for the mean emission rate is $\beta'_m \equiv (e'_m t) = 0.561$. The peak energy E'_0 and the mean energy E'_m are related to the peak and mean emission rates, e'_0 and e'_m, respectively, by Eq. (8).

As originally proposed, transient-current spectroscopy involves recording the emission current over time at fixed temperature and computing the interface-state density from Eq. (18), which reveals that the product $t \times J$ is directly proportional to N_{is}.[11] The energy of the emitting centers at time t is

$$E'_m = kT \ln(\sigma_n v_n N_C t/\beta'_m). \quad (19)$$

An expression similar to Eq. (18) can be derived for isothermal measurements under constant-capacitance conditions. From Eqs. (3), (4), and (7), the time derivative of the gate voltage is related to the interface-state density as follows:

$$dV_G/dt = (A/C_i)qN_{is}(E'_m)\Delta E'_n/t. \quad (20)$$

From Eq. (19) it follows that the isothermal transient response, whether a current or gate-voltage transient, must be monitored over many decades of time in order to acccess a reasonably wide energy interval in the semiconductor bandgap. Alternatively, transient-current spectroscopy can be performed as a DLTS measurement by selecting a specific delay time t_1 in Eqs. (18) and (19) and ramping the temperature as described in previous sections. In transient-current DLTS only a single delay time is required to define an emission rate window. This is a consequence of the current being proportional to the *rate of change* of charge in interface states. In CC-DLTS the gate voltage senses the *net charge* in interface states, and therefore two delay times are required to establish a rate window.

D. Fermi-level probe

For the case of majority-carrier emission after complete trap filling, the energy of the interface states responsible for the DLTS signal must be extracted from the emission rate with Eq. (8). With the assumption that the capture cross section is independent of energy, the interface-state density (versus temperature) is directly obtained from a CC-DLTS spectrum with Eqs. (13) and (14). However, the capture cross section is required to establish the energy scale from Eq. (8), and for characterization of dynamic properties the measurement must yield the capture cross section as well as the interface-state density (versus energy). With defects in crystalline semiconductors, a discrete energy level gives rise to a narrow emission peak in a DLTS spectra which yields the activation energy and capture cross section from an Arrhenius analysis of the emission rate.[7] This approach has also been applied to interface states with the analysis of the previous section.[1,2,14,15,18] However, since the interface-state density is generally smoothly varying, the emission spectrum is essentially featureless, which contributes to imprecise determination of the emission rate at a given (relative) energy and, hence, results in large uncertainties. Alternative methods have been developed for determining the energy of emitting centers independently of the emission rate, which thereby permit evaluation of interface-state distributions with capture cross sections that vary with energy. The basic idea is to use the Fermi level at the interface as a probe for controlling the occupancy of interface traps at known energies. The idea has been implemented by using the amplitude of the charging pulse to position the Fermi level at the interface[19,20] and by selecting the depletion bias to control trap occupancy.[21] The additional information is used to independently obtain the interface-state distribution and capture cross section. The measurement is illustrated here with the *energy-resolved* DLTS technique.[19]

With energy-resolved DLTS, the energy of the traps contributing to the DLTS signal is determined independently of the emission rate by using the amplitude of the charging pulse to selectively populate the interface traps. With reference to Fig. 2, the algebraic sum of the depletion bias and the voltage pulse gives the gate voltage during a charging pulse. For given MIS device and materials parameters, the semiconductor surface potential can be computed from the high-frequency capacitance, measured at a given gate voltage, and used to locate the intersection of the Fermi energy with the semiconductor interface.[6] Interface states situated below the Fermi level are occupied by electrons and those above are empty, the transition occuring over an energy range of $\sim 2kT$ at the Fermi level. Thus, at a fixed temperature, as the pulse amplitude is varied from zero to some maximum value which drives the capacitor into strong accumulation (see Fig. 3), the DLTS signal increases from zero to a saturated value as the Fermi level, during trap filling, passes through that portion of the interface-state distribution which is responsible for the DLTS signal.

The full advantage of the technique is realized by using two adjacent charging pulses of slightly different amplitudes to define a narrow energy interval at the interface in the semiconductor band gap for profiling the interface-state distribution. The gate-voltage waveform is schematically illustrated for an n-type semiconductor in Fig. 6. Voltage pulses are periodically applied to drive the MIS capacitor from depletion toward accumulation. For both pulse amplitudes,

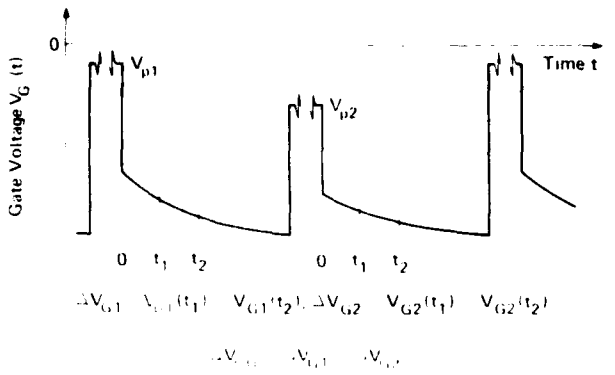


FIG. 6. Schematic diagram of the gate-voltage waveform during an energy-resolved CC DLTS measurement on an MIS device.

the pulse widths are identical and sufficiently long for the capacitor to reach equilibrium during a charging pulse. The same pair of delay times is used after each pulse to provide a common rate window e_0. Since at equilibrium the majority-carrier capture and emission rates are equal, it follows that the pulse width must be greater than the emission time constant, $1/e_0$, to insure equilibrium under given measurement conditions. Separate DLTS signals are recorded for each pulse and subtracted to form a *net* trap-emission signal ΔV_{GG}.

The use of voltage-pulse pairs is the experimental basis of *the double-correlation DLTS technique used to measure* spatial profiles of discrete deep levels in bulk semiconductors.[22-24] Here, the technique is used to measure the *energy* profile of a continuous interface-trap distribution. With reference to the measurement system in Fig. 2, the double-correlation technique is implemented by using a dual-pulse generator to apply pairs of voltage pulses; separate signal integrators are used to record the DLTS signal for each pulse in the pair, and a differential amplifier provides the net emission signal. The supplemental C–V data required for energy-resolved DLTS are readily obtained by using a high-frequency capacitance meter to measure device capacitance as a function of gate voltage at the DLTS-measurement temperature; the results are used to compute the semiconductor surface potential from which is obtained the Fermi energy at the semiconductor–insulator interface.[6]

The energy-resolved DLTS technique is further illustrated in Fig. 7 with an energy-band diagram for the semiconductor–insulator interface in an MIS device which is biased in depletion. The energies F_{p1} and F_{p2} are the Fermi levels corresponding to the gate voltages V_{p1} and V_{p2}, respectively, in Fig. 6. These energies are determined from supplemental C–V measurements and are the levels to which the interface states are populated with electrons during the two different voltage pulses. The mean Fermi level may be defined as follows: $\langle F_p \rangle \equiv (F_{p1} + F_{p2})/2$, and the energy separation between the Fermi levels is $\Delta F_p \equiv |F_{p2} - F_{p1}|$. The net emission signal is sensitive only to interface states located in between F_{p1} and F_{p2}; there is no contribution from states above F_{p1} since they are not populated, and the contribution from states below F_{p2} is cancelled in forming the net emission signal. The signal is a maximum when the interval

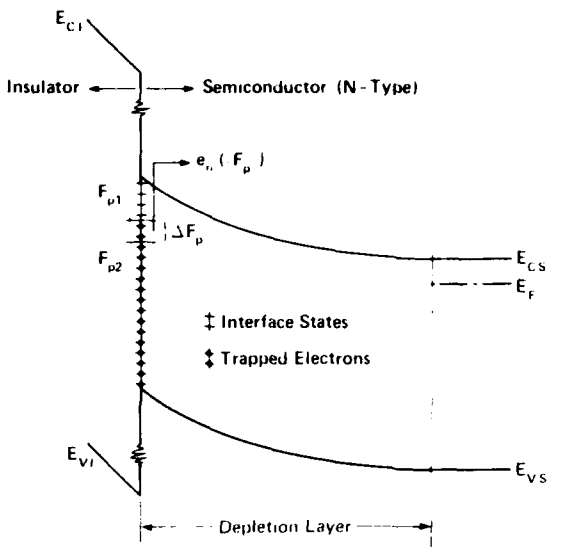


FIG. 7. Energy-band diagram for an MIS device in an energy-resolved DLTS measurement.

bounded by F_{p1} and F_{p2} overlaps that interval of the trap distribution in which $e_n = e_0$. The energy selectivity of the energy-resolved DLTS technique is achieved with small values for ΔF_p; at maximum signal $\langle F_p \rangle = E_0$ in the limit as $\Delta F_p \to 0$; for a broad Fermi–Dirac difference distribution (i.e., with F_{p1} and F_{p2} widely spaced) $\langle F_p \rangle$ approaches E_m at maximum signal. As discussed in Sec. III C 1 the difference between E_0 and E_m is generally less than kT and can be neglected.

Energy-resolved DLTS is analytically formulated as follows: For electron emission from a continuous distribution of interface traps, the net emission signal under constant-capacitance conditions is

$$\Delta V_{GG} = (A/C_i) \int_{-\infty}^{\infty} qN_{is}(E) \times [\exp(-e_n t_1) - \exp(-e_n t_2)](F_1 - F_2)dE, \quad (21)$$

with

$$F_j \equiv \{1 + g_{is} \exp[(F_{pj} - E)/kT]\}^{-1}.$$

Here, F_j is the Fermi–Dirac distribution function for electron occupation of interface states, and F_{pj} is measured relative to the semiconductor conduction-band minimum at the interface. As noted previously, it is necessary to assume a value of unity for the spin degeneracy factor g_{is}. In the integrand of Eq. (21) the difference of the Fermi–Dirac distributions yields a function which is symmetrically distributed about the energy $\langle F_p \rangle$. For closely spaced Fermi levels (i.e., $\Delta F_p \leq 30$ meV) the function is sharply peaked, and N_{is} may be treated as a constant with magnitude $N_{is}(\langle F_p \rangle)$. Hence, the integral is the cross-correlation function between the trap-emission distribution, $\exp(-e_n t_1) - \exp(-e_n t_2)$, and the Fermi–Dirac difference distribution, with the Fermi level (e.g., F_{p2} with ΔF_p fixed) serving as the scanning parameter. These two distributions are shown in Fig. 8 for hypothetical interface-trap parameters and experimental conditions. On the left is the trap-emission distribution for

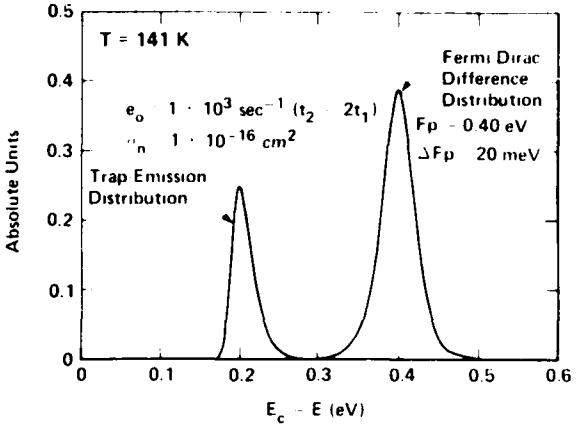


FIG. 8. Illustrations of the trap-emission distribution and the Fermi–Dirac difference distribution for energy-resolved DLTS with hypothetical interface-trap parameters and experimental conditions.

traps emitting electrons in the rate window e_0. On the right is the Fermi–Dirac difference distribution for the specific case of $\langle F_p \rangle = 0.40$ eV and $\Delta F_p = 20$ meV. From Eq. (21) the net emission signal is proportional to the integral of the product of the above two distributions, that is, to the cross-correlation function of the two distributions. The difference distribution is translated in energy, at a fixed temperature, by varying the pulse amplitudes. Then, the cross-correlation function is a maximum when $e_n(\langle F_p \rangle) = e_0$. From this the capture cross section $\sigma_n(\langle F_p \rangle)$ can be computed with Eq. (8). Performing the measurement over a range of fixed temperatures or with different values of e_0 yields σ_n at different energies.

The analyses of this and the previous section will be illustrated with a measurement of electron traps at the Si–SiO_2 interface. The MOS capacitor consisted of an n-type silicon substrate with a 115-nm, thermally grown oxide layer and an aluminum gate electrode. In Fig. 9(a) the mean Fermi energy $\langle F_p \rangle$ for maximum ΔV_{GG} is shown for several measurement temperatures. The electron capture cross section was computed for each measurement temperature from Eq. (8) and is plotted in Fig. 9(b). The large error bars reflect the exponential dependence of σ_n on E in Eq. (8). Within the experimental uncertainty, σ_n is a constant. In cases where σ_n is found to vary significantly, measurements over a range of emission rate windows e_0 at each temperature would be used to determine temperature versus energy dependences.

With σ_n a known constant, the single CC–DLTS emission spectrum shown in Fig. 10, obtained with the saturating pulse technique, can be directly analyzed with Eqs. (8), (13), and (14) to obtain $N_{ss}(E)$. Equation (8) is used to convert the temperature scale of the emission spectrum to an energy scale in the semiconductor bandgap, and Eqs. (13) and (14) yield the interface-state density. The resulting distribution is also shown in Fig. 10; the interface-state density increases monotonically with energy from midgap toward the silicon conduction band. The dashed line segment in the interface-state distribution is explained in the next section.

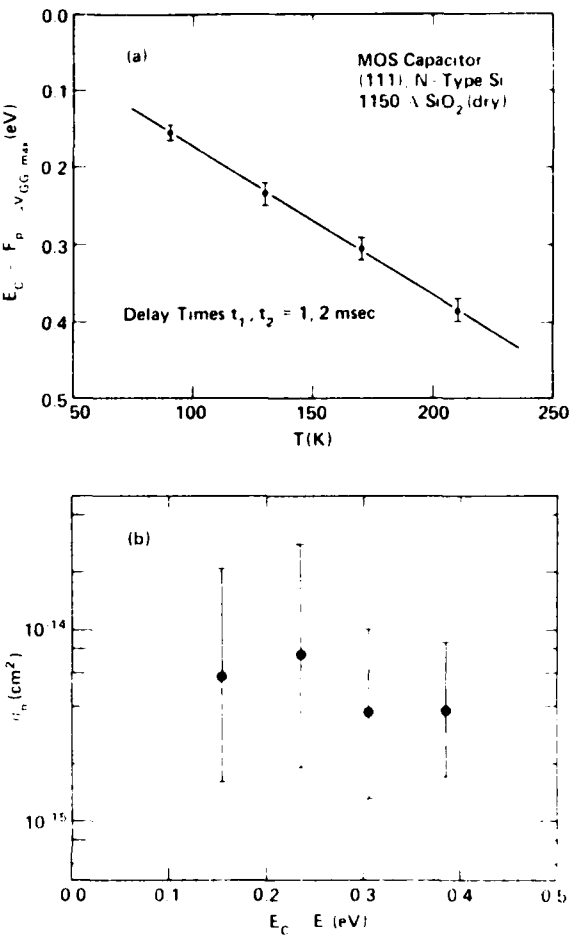


FIG. 9. Results for energy-resolved DLTS measurements of electron traps at the Si–SiO_2 interface: (a) mean Fermi energy $\langle F_p \rangle$ for maximum net emission signal $\Delta V_{GG\,max}$ vs temperature and (b) electron capture cross section σ_n vs energy in the silicon bandgap.

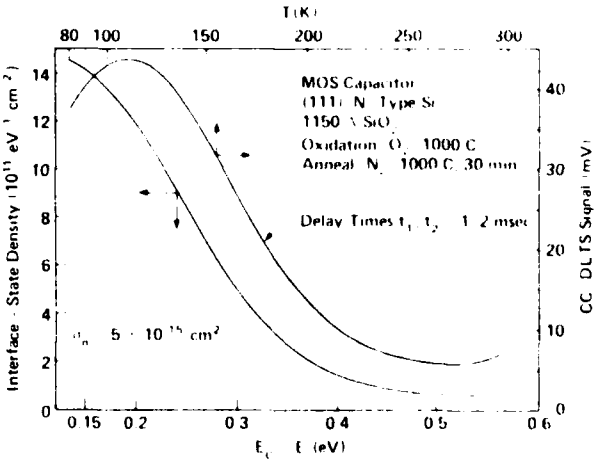


FIG. 10. CC-DLTS emission spectrum and computed interface-state distribution with σ_n = 5 · 10^{-15} cm^2.

E. Surface generation

The analysis developed in Sec. III C is applicable when the CC–DLTS signal is dominated by majority-carrier emission. However, near the semiconductor midgap, minority-carrier emission from interface states becomes a competitive process, and a complete analysis of the entire recovery to equilibrium, after trap filling, must include emission and capture of both charge carriers. With the onset of significant minority-carrier emission from interface states, surface generation contributes to the DLTS signal in MIS capacitors. This is a consequence of surface generation contributing to the change in interface charge with time through the emission of majority carriers and the inability of emitted minority carriers to escape from the interface.

The effect of surface generation is illustrated in Fig. 11 with CC–DLTS spectra obtained from two differently processed MOS capacitors on p-type silicon. Each spectrum displays a prominent peak above room temperature, which corresponds to emission from energies near midgap. The interface-state densities for these devices, as measured by the quasistatic capacitance–voltage technique, are essentially constant over the midgap region.[1] Furthermore, the position and shape of these peaks strongly depend on the depletion bias during the CC–DLTS measurement. The appearance of such peaks and their dependence on experimental conditions is to be expected from the participation of minority carriers in the transient response.

A simplified formulation demonstrates how minority-carrier effects can cause the high-temperature peaks in Fig. 11. The primary mechanism is surface generation, although capture and emission of both charge carriers are required to describe the entire relaxation to equilibrium. With the inclusion of emission and capture of electrons and holes in an n-type semiconductor, the energy interval in Eq. (13) can be formally expressed as follows:

$$\Delta E_\Sigma = \int [(e_n + c_p)/e_\Sigma] \times [\exp(-e_\Sigma t_1) - \exp(-e_\Sigma t_2)]\, dE, \tag{22}$$

where $e_n(e_p)$ and $c_n(c_p)$ are the rate coefficients for emission and capture, respectively, of majority (minority) carriers and $e_\Sigma \equiv e_n + e_p + c_n + c_p$. The capture and emission rate coefficients for electrons are defined in Eqs. (6) and (8), respectively. The corresponding rate coefficients for holes are

$$e_p = \sigma_p v_p N_V \exp[-(E_G - E)/kT] \tag{23}$$

and

$$c_p = \sigma_p v_p p_s, \tag{24}$$

where σ_p is the capture cross section for holes, v_p is the mean thermal velocity for holes, N_V is the effective density of states in the semiconductor valence band, E_G is the width of the semiconductor bandgap, and p_s is the free hole concentration at the semiconductor–insulator interface. In Eq. (22) the integration is over the semiconductor bandgap. The formulation focuses on the effect of surface generation and specifically neglects the time dependence of p_s in Eq. (24).

In Fig. 12 is shown the computed variation of e_Σ with energy in silicon near midgap for selected temperatures.

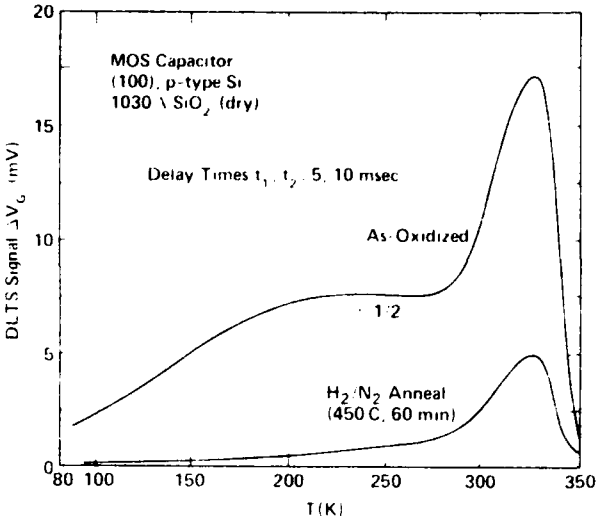


FIG. 11. CC–DLTS spectra for a silicon MOS capacitor as-oxidized and after a low-temperature anneal.

Superimposed on this family of curves is a representative emission rate window e_0. The intersections of the rate window with a given curve determine the energy interval over which interface traps contribute to the DLTS signal. At low temperatures (e.g., 225 K) the signal is dominated by e_n and the energy interval is ΔE_n. As the temperature increases, minority carriers begin to contribute significantly (e.g., at 295 K) and the energy interval increases beyond ΔE_n. With further increase in temperature the energy interval rapidly decreases to zero (e.g., at 340 K) as e_Σ becomes greater than e_0 at all energies. The dependence on experimental conditions (i.e., the gate voltage in depletion) results from the depletion-bias-determined intersection of the Fermi level with the interface which specifies the equilibrium occupation of the interface states; states below the Fermi level do not

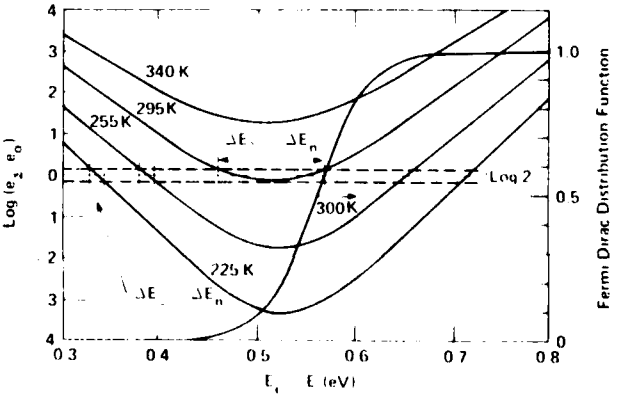


FIG. 12. Variation of e_Σ, which is the sum of the majority and minority carrier emission and capture coefficients (i.e., $e_\Sigma = e_n + e_p + c_n + c_p$), with energy near the silicon midgap for selected temperatures. The capture cross sections used in the calculation are $\sigma_n = 1 \times 10^{-15}$ cm^2 and $\sigma_p = 1 \times 10^{-14}$ cm^2, and the emission rate window is $e_0 = 1 \times 10^3$ s^{-1} ($t_2 = 2t_1$). Also shown is the Fermi–Dirac distribution function at 300 C with the Fermi level intersecting the interface at midgap.

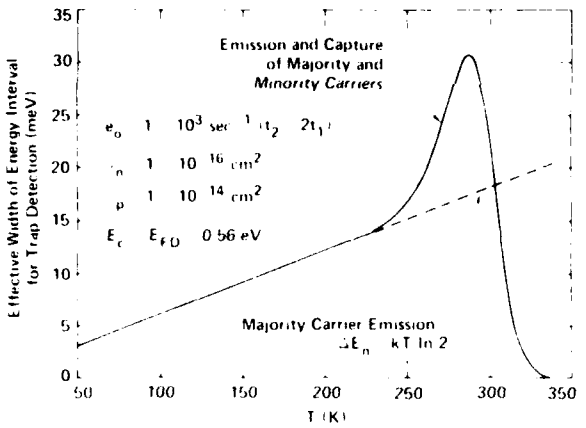


FIG. 13. Temperature dependence of the energy interval ΔE_Σ for detecting interface states in a CC–DLTS measurement. The solid line includes emission and capture of majority carriers, and the dashed line is for majority-carrier emission only.

change occupancy during the transient response and therefore cannot contribute to the DLTS signal.

The effect of surface generation is further illustrated in Fig. 13 which shows the dependence of ΔE_Σ on temperature as predicted by Eq. (22), for representative values of the capture cross sections and with the assignment of equilibrium values for the capture coefficients which correspond to the Fermi level intersecting the interface at midgap. At low temperatures, $\Delta E_\Sigma = \Delta E_n$. However, at high temperatures ΔE_Σ sharply departs from the linear temperature dependence of majority-carrier emission. From Eq. (13) it follows that such structure in ΔE_Σ can produce an apparent emission peak in a DLTS spectrum (as in Fig. 11) and a corresponding artifactual peak in the interface-state distribution. Thus, interface-state distributions computed with the majority-carrier analysis are generally not accurate near the semiconductor midgap; this is the signficance of the dashed line segment in the interface-state distribution in Fig. 10. On the other hand, the model presented here can be applied to estimate the cross section of interface states for capturing minority carriers in MIS capacitors.

IV. DISCUSSION AND CONCLUSIONS

In the limit of low interface-state densities, either the transient-capacitance or the constant-capacitance mode (or current-transient DLTS) can be used for accurate measurement. A practical maximum sensitivity for either mode can be estimated from the generally observed capability of DLTS spectrometers for detecting deep levels in bulk semiconductors. It is commonly found that the minimum deep level density is $\sim 10^{-4}$ of the shallow dopant concentration. Applied to an MIS device, this is approximately the ratio of the interface-charge density and the semiconductor space-charge density. The latter can be computed from Eq. (A4) for given device parameters and gate voltage. For example, a silicon MOS capacitor, with a 100-nm oxide and a dopant concentration of 1×10^{15} cm^{-3}, has a silicon space-charge density of $\sim 7\times10^{10}$ cm^{-2} at room temperature with the Fermi level at the interface situated near midgap. From Eqs. (13) and (14), with $t_2/t_1 = 2$, the above sensitivity factor yields a minimum detectable interface-state density of $\sim 4\times10^{8}$ eV^{-1} cm^{-2}.

Features of the different modes of measurement can be compared as the interface-state density increases. The restrictive upper limit on trap density for the applicability of the transient-capacitance mode as expressed by Eq. (A7) is well satisfied only with state-of-the-art thermally grown oxides on single-crystal silicon. As discussed in the appendix, calculating the interface-state density from transient-capacitance data requires both the doping concentration and the high-frequency capacitance for a given depletion bias. Since the high-frequency capacitance at a fixed gate bias changes with temperature and at high trap densities varies by a significant fraction of the equilibrium value during the transient response (see Appendix), the prefactor relating the interface-state density to the DLTS signal, Eqs. (A6) and (A8), also changes with time and temperature which, if not accounted for, will yield erroneous interface-state distributions. Transient-current DLTS is not limited to low interface-state densities but the proportionality factor in Eq. (18) does contain the high-frequency capacitance. From Eq. (13) in the constant-capacitance mode the proportionality factor depends only on the insulator capacitance per unit area.

The use of the Fermi level to control interface-trap occupancy is a unique property of the MIS structure, which results from the dependence of the semiconductor surface potential on gate bias. A feature of the Fermi-level probe in DLTS measurements, as compared to the conventional Arrhenius analysis, is the use of the Fermi energy to identify the free energy for thermal emission and to compute the capture cross section. An Arrhenius analysis of the emission rate yields the activation energy for thermal emission, which is an enthalpy, and only the product of the capture cross section and a factor dependent on the change in entropy accompanying charge emission.[25,26] However, one key feature of majority-carrier emission after complete trap filling is not retained with the Fermi-level probe. In the former case, the DLTS emission signal is not directly affected by fluctuations in the semiconductor surface potential since all interface traps are populated during the voltage pulse and the transient response depends only on the emission rate, Eq. (8). In MOS devices consisting of thermally grown SiO_2 on Si, such fluctuations arise from the random spatial distribution of fixed positive charge in the oxide and have been used to explain results from conductance measurements of interface states.[5,27] With the Fermi-level probe, the occupation of detected interface traps is determined by a Fermi level and hence the DLTS signal is affected by surface-potential fluctuations.[19,21,28] In energy-resolved DLTS this effectively broadens the energy interval ΔF_p, which reduces the resolution with which the energy interval of detected traps can be measured.

The principal simplification in the analyses of Sec. III is neglect of bulk deep levels in the semiconductor depletion layer. The emission spectrum for such defects is superimposed on that arising from interface states. In single-crystal semiconductors, bulk deep levels appear at discrete energies and, therefore, introduce relatively sharp peaks compared to the nearly structureless spectrum which usually arises from a continuum of states. In crytalline silicon, the bulk defect

concentration is generally negligible compared to the density of interface states; silicon MOS devices, with low interface-state densities, have been used to study bulk defects which were introduced by ion implantation.[9] In the energy-resolved DLTS technique for interface-state measurement (Sec. III D), the dual-pulse feature minimizes any contribution from bulk deep levels because the bulk contribution is largely cancelled by forming the difference of two separate DLTS signals (Figs. 6 and 7). This is exemplified by the use of the dual-pulse technique on current rectifying devices (i.e., Schottky-barrier and *p-n* junction diodes) to achieve high spatial resolution for defect depth profiling.[27,28] Similarly, in DLTS measurements on MIS devices the pulse or depletion bias can be varied in order to modulate the relative contributions to the emission signal from bulk and interface defects.[28,29] This relies on the total number of bulk defects which contribute to the signal varying with the semiconductor depletion width while the contribution from interface states is independent of depletion width. An alternative approach can be outlined which eases the analytical complexity of deconvolving superimposed bulk and interface emission spectra. An MIS capacitor from which a composite spectrum has been recorded, can be reprocessed into a Schottky-barrier diode. This device can then be used for DLTS measurement of the bulk defects without interference from interface states. In general, the analysis will entail measuring the spatial profiles of the bulk deep levels and extrapolating the bulk defect distributions to the interface for deconvolving the MIS emission spectrum. The above considerations underscore the versatility of DLTS for characterizing both interface and bulk defects in semiconductors.

In conclusion, deep-level transient spectroscopy is a particularly powerful diagnostic technique for the characterization of electronic defect levels at a semiconductor–insulator interface. It features high sensitivity ($<1\times10^{9}\,\mathrm{eV}^{-1}\,\mathrm{cm}^{-2}$), high energy resolution ($\sim kT$), determination of dynamic properties (i.e., capture cross sections), and versatility (applicable to both bulk and interface defect evaluation). The full potential of DLTS for MIS interface studies is realized in the constant-capacitance mode of measurement.

ACKNOWLEDGMENTS

The author is pleased to acknowledge helpful discussions with M. Schulz and J. M. Brazio. The work was supported by the U. S. Army ERADCOM.

V. APPENDIX

A. Transient-capacitance DLTS

For transient-capacitance measurements on MIS capacitors, the device is biased in depletion with a fixed gate voltage and periodically pulsed into accumulation to populate all interface states with majority carriers. The DLTS signal $\Delta C_{\rm hf}$ is obtained as the difference of the high-frequency device capacitance $C_{\rm hf}$ measured at times t_1 and t_2 after the trap-filling pulse

$$\Delta C_{\rm hf}\equiv C_{\rm hf}(t_1)-C_{\rm hf}(t_2). \tag{A1}$$

To obtain an analytical expression for $\Delta C_{\rm hf}$ consider an MIS capacitor on an *n*-type uniformly doped semiconductor which possesses interface states and a nonzero metal–semiconductor work function difference $\Phi_{\rm MS}$. The device is otherwise ideal, and the work function difference is included because it varies with temperature and therefore can be significant in a DLTS temperature scan. In addition, the shallow donors, of concentration N_D, are assumed to be completely ionized over the temperature range of interest (e.g., $\gtrsim 100$ K for silicon). For a depletion bias, the high-frequency capacitance is the series combination of the insulator capacitance C_i and the semiconductor depletion capacitance C_D. The gate voltage develops a potential difference across the insulator V_i and a surface potential in the semiconductor ψ_s as follows:

$$V_G = V_i + \psi_s + \Phi_{\rm MS}/q. \tag{A2}$$

From Gauss' law, the potential difference across the insulator depends on the semiconductor space charge Q_s (per unit area) and the net charge in interface states Q_{is} (per unit area) as follows:

$$V_i = -A(Q_s + Q_{is})/C_i, \tag{A3}$$

where A is the gate area. The semiconductor space charge, due to ionized dopant, is given by the following expression accurate for depletion[4,6]:

$$Q_s = 2\epsilon_s kT\{[q\psi_s/(kT)] - 1\}^{1/2}/(qL_D), \tag{A4}$$

where L_D is the extrinsic Debye length,

$$L_D\equiv[2\epsilon_s kT/(q^2N_D)]^{1/2}.$$

From Eqs. (A1)–(A4) and the definition of the semiconductor depletion capacitance, $C_D\equiv\partial Q_s/\partial\psi_s$, the following expression for $C_{\rm hf}$ in depletion is derived:

$$C_{\rm hf}/C_i = \{1-[2C_i^2/(A^2q\epsilon_sN_D)]\times[V_G-(\Phi_{\rm MS}/q)+(kT/q)+(AQ_{is}/C_i)]\}^{-1/2}, \tag{A5}$$

where $V_G-(\Phi_{\rm MS}/q)+(kT/q)+(AQ_{is}/C_i)<0$ for depletion on an *n*-type semiconductor.

Separation of the interface-state charge density into equilibrium and time-varying components permits evaluation of the conditions under which the DLTS signal is linearly proportional to the difference in interface-state charge measured at times t_1 and t_2 after a trap-filling pulse. In equilibrium, the depletion bias fixes the Fermi level near the middle of the semiconductor bandgap at the interface; the corresponding equilibrium interface charge density will be designated $Q_{is}(\infty)$. After the gate voltage is pulsed toward accumulation, the nonequilibrium interface charge $Q_{is}(t)$ relaxes with time to $Q_{is}(\infty)$. The difference between these interface-state charge densities is $\Delta Q_{is}(t)\equiv Q_{is}(t)-Q_{is}(\infty)$. With $\Delta Q_{is}(t)+Q_{is}(\infty)$ substituted for Q_{is}, Eq. (A5) can be expanded in ΔQ_{is} with retention of terms through first order and substituted into Eq. (A1) to obtain

$$\Delta C_{\rm hf} = [C_{\rm hf}^3/(Aq\epsilon_sN_DC_i)][Q_{is}(t_1)-Q_{is}(t_2)]. \tag{A6}$$

To neglect the higher order terms in the expansion of Eq. (A5), the following inequality must be satisfied:

$$|\Delta Q_{is}(t_1)| \ll [(Aq\epsilon_sN_D)/(2C_i)] - (C_i/A)\{V_G-(\Phi_{\rm MS}/q)+(kT/q)+[AQ_{is}(\infty)/C_i]\}, \tag{A7}$$

where the expression bracketed by $\{\ \}$ is negative for depletion on an n-type semiconductor. Finally, from the discussion in Sec. III C, for a constant capture cross section the interface-state density is related to the DLTS signal as follows:

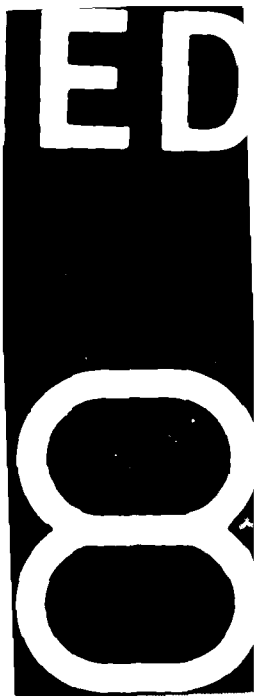
$$\Delta C_{\rm hf} = f_C q N_{\rm ss}(E_m) kT \ln(t_2/t_1), \tag{A8}$$

where f_C is the proportionality factor in Eq. (A6).

Equation (A8) has been used to analyze interface states in transient-capacitance DLTS measurements.[10,14,18] However, the derivation presented here permits an evaluation of the limits of applicability of the transient-capacitance mode. Specifically, Eq. (A7) must be satisfied.[30] To estimate the right-hand side of the inequality, consider the representative case of a silicon MOS capacitor biased in depletion with a 100-nm oxide layer and $N_D = 1\times10^{15}\ {\rm cm}^{-3}$. Equation (A7) then requires that $\Delta Q_{\rm ss}(t_1)/q \ll 1\times10^{11}\ {\rm cm}^{-2}$. Since $\Delta Q_{\rm ss}(t_1)$ represents the change in interface charge due to trap emission in essentially the upper half of the bandgap, Eq. (A7) requires that the *average* interface-state density be much less than $\sim 2\times10^{11}\ {\rm eV}^{-1}\ {\rm cm}^{-2}$. Interfaces with defect densities substantially less than this (e.g., $\lesssim 5\times10^{10}\ {\rm eV}^{-1}\ {\rm cm}^{-2}$) are readily achieved only with state-of-the-art thermally grown oxides on crystalline silicon.

From Eqs. (A6) and (A8) it follows that calculation of the interface-state density from transient-capacitance data requires the doping concentration and the high-frequency capacitance for the depletion bias. Provided that Eq. (A7) is well satisfied, the proportionality factor f_C is essentially constant at a given temperature. However, from Eq. (A5) f_C is implicitly temperature dependent. This is evident from the observation that the $C_{\rm hf} - V_G$ characteristic of an MIS device shifts with temperature (see for example Refs. 31 and 32). This shift is essentially parallel to the voltage axis and results in the high-frequency capacitance, and therefore f_C, at a fixed depletion bias changing with temperature. With the transient-capacitance mode of measurement, this variation can be incorporated in the data analysis with a supplemental measurement of the temperature dependence of $C_{\rm hf}(V_G)$. Alternatively, V_G can be adjusted to maintain a constant equilibrium value of $C_{\rm hf}$ as the temperature is scanned to record $\Delta C_{\rm hf}(T)$.[33] This latter approach is automatically implemented with constant-capacitance DLTS.

[1] N. M. Johnson, D. J. Bartelink, and M. Schulz, in *The Physics of SiO_2 and its Interfaces*, edited by S. T. Pantelides (Pergamon, New York, 1978), pp. 421–427.
[2] M. Schulz and E. Klausmann, Appl. Phys. **18**, 169 (1979).
[3] A. S. Grove, *Physics and Technology of Semiconductor Devices* (Wiley, New York, 1967).
[4] S. M. Sze, *Physics of Semiconductor Devices* (Wiley, New York, 1981, 2nd ed.
[5] E. H. Nicollian and J. R. Brews, *MOS Physics and Technology* (Wiley, New York, 1982).
[6] A. Goetzberger, E. Klausmann, and M. Schulz, CRC Rev. Solid State **6**, 1 (1976).
[7] D. V. Lang, J. Appl. Phys. **45**, 3014 (1974); **45**, 3023 (1974).
[8] G. L. Miller, D. V. Lang, and L. C. Kimerling, in *1977 Annual Review of Material Science* (Annual Reviews, Palo Alto, 1977), pp. 337–448.
[9] K. L. Wang and A. O. Evwaraye, J. Appl. Phys. **47**, 4574 (1976).
[10] K. L. Wang, IEEE Trans. Electron Devices **ED-27**, 2231 (1980).
[11] T. G. Simmons and L. S. Wei, Solid-State Electron. **17**, 117 (1974).
[12] N. M. Johnson, D. K. Biegelsen, and M. D. Moyer, J. Vac. Sci. Technol. **19**, 390 (1981).
[13] M. D. Jack, R. C. Pack, and J. Henriksen, IEEE Trans. Electron Devices **ED-27**, 2226 (1980).
[14] M. Schulz and N. M. Johnson, Solid-State Commun. **25**, 481 (1978); Errata **26**, 1261 (1978).
[15] E. Klausmann, Inst. Phys. Conf. Ser. **50**, 97 (1980).
[16] H. Deuling, E. Klausmann, and A. Goetzberger, Solid-State Electron. **15**, 559 (1972).
[17] R. J. Kriegler, T. F. Devenyi, K. D. Chik, and J. Shappir, J. Appl. Phys. **50**, 398 (1979).
[18] M. Schulz and N. M. Johnson, Appl. Phys. Lett. **31**, 622 (1977).
[19] N. M. Johnson, Appl. Phys. Lett. **34**, 802 (1979).
[20] K. L. Wang, IEEE Trans. Electron Devices **ED-26**, 819 (1979).
[21] T. J. Tredwell and C. R. Viswanathan, Solid-State Electron. **23**, 1171 (1980).
[22] H. Lefevre and M. Schulz, Appl. Phys. **12**, 45 (1977).
[23] N. M. Johnson, D. J. Bartelink, R. B. Gold, and J. F. Gibbons, J. Appl. Phys. **50**, 4828 (1979).
[24] C. R. Crowell and S. Alipanahi, Solid-State Electron. **24**, 25 (1981).
[25] J. A. Van Vechten and C. D. Thurmond, Phys. Rev. B **14**, 3539 (1976).
[26] O. Engstrom and A. Alm, Solid-State Electron. **21**, 1571 (1978).
[27] E. H. Nicollian and A. Goetzberger, Bell Syst. Tech. J. **46**, 1055 (1967).
[28] E. Klausmann, in *Insulating Films on Semiconductors*, edited by M. Schulz and G. Pensl (Springer, New York, 1981), pp. 169–173.
[29] K. Yamasaki, M. Yoshida, and T. Sugano, Jap. J. Appl. Phys. **18**, 113 (1979).
[30] If the capacitance transient is recorded over time at fixed temperature, as in Sec. III C 2, the interface-state density may be computed from $d(1/C_{\rm hf}^2)/dt = [2/(Aq\epsilon_s N_D C_i)]q N_{\rm ss}(E_m')\Delta E_m'/t$. This expression is derived directly from Eq. (A5) and therefore is not restricted in application by Eq. (A7).
[31] P. V. Gray and D. M. Brown, Appl. Phys. Lett. **8**, 31 (1966).
[32] M. R. Boudry, Appl. Phys. Lett. **22**, 530 (1973).
[33] E. Rosencher, A. Chantre, and D. Bois, in *Physics of MOS Insulators*, edited by G. Lucovsky, S. T. Pantelides, and F. L. Galeener (Pergamon, New York, 1980), pp. 331–335.


ED